Measuring Benefits from New Products in Markets with Information Frictions

Ilya Morozov[*]
Northwestern University

October 9, 2022

Abstract

I study how much consumers benefit from new products in markets with information frictions. I analyze new products in the U.S. hard drive market, which is characterized by ample product innovation. Using unique click-stream data, I measure the magnitude of two frictions, category consideration and costly search, and show that both play a crucial role in shaping consumer demand. To estimate consumer surplus from new products, I develop a search model that captures both frictions and propose a novel Bayesian estimation method to recover its parameters. I then show that ignoring information frictions leads researchers to underestimate the consumer surplus from new hard drives because it appears that consumers do not value the combinations of attributes these hard drives offer. Partly eliminating frictions, through marketing efforts or market-wide transparency initiatives, can help consumers to more fully internalize the benefits of new product launches.

Keywords: New Products, Value of Innovation, Consumer Search, Bayesian Analysis.
JEL codes: C11, D83, O33.

[*]I am grateful to my advisors, Bart Bronnenberg, Peter Reiss, and Stephan Seiler, for their exceptional supervision and constant support. I also thank a group of researchers whose insightful comments and suggestions have greatly improved the paper: Paulo Albuquerque, Kristina Brecko, Robert Donnelly, Liran Einav, Pedro Gardete, Brett Gordon, Wesley Hartmann, Avery Haviv, Han Hong, Megan Hunter, Przemyslaw Jeziorski, Anja Lambrecht, Suraj Malladi, Navid Mojir, Sridhar Moorthy, Sridhar Narayanan, Andres Perlroth, Peter Rossi, Navdeep Sahni, Paulo Somaini, Raluca Ursu, Xavier Warnes, and Ali Yurukoglu, as well as seminar participants at Stanford University, INSEAD, Tilburg University, London Business School, Harvard University, Northwestern University, University of Washington, Rochester University, New York University, UC Berkeley, and UCLA. I am greatly indebted to Julie Williamsen, Sean Kaneshiro, and Wonhee Lee for helping me access Comscore data. I thank Robyn Lockwood for teaching me to write well, although it will take a lifetime to master what she taught me. All errors are my own.

1 Introduction

Many industries grow through a constant stream of technological innovations that allow firms to improve existing products or introduce entirely new goods. But consumers often struggle to stay informed about new product launches, especially in markets with large assortments. To break this information barrier, firms invest millions of dollars in advertising new products and promoting the value these products create.[1] While industry practitioners see such investments as key to launching new products, the academic literature on this topic has adopted a more simplified view. Much of this literature assumes that consumers become immediately aware of all new products and learn their attributes at no cost (Hausman, 1996; Petrin, 2002). Even when consumers do not adopt a new product due to their limited knowledge, this standard approach ascribes new product failures to consumer preferences. By mistaking information for preferences, this approach may obscure the true value of new products for consumers. It may also prevent researchers from studying how firms' marketing efforts can help consumers fully internalize the benefits of new product launches.

This paper has two goals. First, I extend the existing techniques for estimating the value of new goods by accounting for information frictions. I model two specific frictions, *category consideration* and *costly search*, both of which are grounded in empirical evidence. A consumer first chooses which product category to consider within a specific market. By eliminating certain categories, a consumer might not even consider new products regardless of their attributes. In practice, such behavior may arise because a consumer is unaware of certain product types, finds some categories too complex to understand, or chooses to reduce the consideration set in order to simplify the choice process. Having chosen categories to consider, the consumer learns the basic attributes of all products within these categories but faces residual uncertainty about products' match values. The consumer then examines products one by one to gather missing information, the process I capture using the sequential search model of Weitzman (1979). Since search is costly, the consumer only examines new products if their known attributes seem sufficiently appealing. Importantly, this model allows for the possibility that many consumers do not purchase new products despite the benefits these products offer.

Second, I apply the model to study consumer surplus from new products in the U.S. hard drive market, which is characterized by fast-paced product innovation. I estimate the model using detailed click-stream data from the Comscore Web Behavior Panel, and I measure the extent to which consumers benefited from the introduction of solid state drives (SSDs) – a new class of hard drives that substantially increased the read and write speeds of traditional hard disk drives (HDDs). Consumers did not immediately recognize the benefits of SSDs. A survey conducted by Western Digital in 2016, a decade after the commercial introduction of SSDs, revealed that over 40% of U.S. consumers were still unaware of the solid state drive technology or could not explain how it differs

[1] For example, Toyota spent more than \$1.5 billion on advertising newly released *Corolla*, *Camry*, and *Prius* models in 2019. Disney spent more than \$200 million in 2019 to advertise its new *Avengers: Endgame* movie. Apple spent \$132.2 million on TV and online ads in September and October 2019 to promote the new *iPhone 11* and *Apple TV* Plus service. Source: iSpot and Kantar AdSpender databases.

from HDDs.[2] Therefore, it is possible that some consumers did not even consider the category of SSDs when buying a hard drive. In addition, because of high product turnover and large product assortments, consumers in this market faced a non-trivial search problem. The question I ask in this context is whether and to what extent information frictions affected demand for SSDs, and if so, how that affected the surplus consumers derived from the introduction of SSDs.

The central empirical challenge I need to address is how to separate consumer preferences from information frictions. That consumers do not click on SSDs might be evidence that they have decided not to learn the attributes of these hard drives or chose not to consider the SSD category altogether. It might also reflect that consumers do not like the combinations of attributes that SSDs offer. To measure search frictions, I use a unique click-stream dataset from Comscore that contains the universe of web pages visited by consumers in 2016. This URL-level dataset is substantially more detailed than domain-level Comscore data used in prior work (De Los Santos et al., 2012; De los Santos, 2018). Importantly, from these data, I know how many and which hard drives each consumer examined before buying a hard drive online. By combining this direct measure of search with data on the actual online purchases, I can empirically separate search costs from consumers' preferences for hard drive attributes (e.g., price, storage capacity, and speed).

A further challenge is how to distinguish consideration from search, as both frictions may prevent consumers from buying SSDs. To separate one from another, I rely on two sets of exclusion restrictions. The first one excludes prices from the category consideration stage but allows consumers to react to price changes when choosing what products to search. That is, a consumer who does not consider SSDs does not know their prices and therefore cannot possibly react when these prices change. Asymmetric reaction to temporary price promotions of SSDs and HDDs then allows me to separate consideration from costly search. The second exclusion restriction builds on novel variables I construct from consumer-level data. For example, I identify consumers who visit informational websites about PC hardware. I then use visits to such websites as a consideration shifter, implying that these consumers might be more likely to consider SSDs due to their expertise in computer hardware. Similarly, I construct a variable capturing how many products consumers search in other product categories (i.e., other than hard drives). Because such a variable partly captures individual differences in the opportunity cost of time, I use it as a search cost shifter. These variables generate additional exclusion restrictions, helping me separate consideration from search.

It is generally difficult to estimate structural choice models with search frictions. The prior work shows that traditional frequentist methods of estimating search models might be slow and numerically unstable (Chung et al., 2018). Estimation becomes even more challenging in my model where I add a consideration stage and allow for rich heterogeneity in preferences, search costs, and consideration probabilities. To overcome this issue, I develop a novel Bayesian estimator that incorporates category consideration and costly search into the standard hierarchical probit choice

[2] Western Digital surveyed several thousand participants in the U.S., Spain, Germany, France, and the UK through online surveys and on-site interviews at electronics stores (Western Digital, 2016).

model (Rossi et al., 2012, p.75). Compared to frequentist methods, the MCMC sampler I propose is more numerically stable, can more easily handle rich consumer heterogeneity, and scales better to large assortments commonly observed in online markets.

My estimation results show that introducing SSDs raised the surplus of the average consumer by \$3.2, about 3% of the average hard drive price. By contrast, a perfect information model, similar to those used in the prior work on new products, underestimates this surplus change almost by a factor of three. This bias is driven by two distinct effects. On the one hand, the perfect information model inflates the surplus change by incorrectly assuming that consumers know all attributes of SSDs and therefore always know whether some SSDs match their preferences better than HDDs. This is not the case in my model where many consumers do not consider SSDs or remain uninformed about SSDs' attributes. On the other hand, the perfect information model underestimates the surplus change by incorrectly attributing the low market share of SSDs to consumer preferences. The net result of these two opposing effects is that the perfect information model dramatically underestimates the surplus generated by SSDs. Accounting for information frictions, therefore, is crucial for estimating consumer surplus from new products in this market.

Modeling frictions also helps understand how reducing these frictions can affect the surplus generated by new products. My estimates reveal that the magnitude of information frictions substantially affects how much consumers benefit from SSDs. The surplus from SSDs increases by 60% when I partly remove frictions from the estimated model while keeping preferences fixed. In particular, increasing the number of tech-savvy users who frequently consider SSDs raises the surplus from SSDs from \$3.2 to \$4.8. Reducing search costs of all consumers by about 50% further raises this surplus change to \$5.0. This observation suggests that one can help consumers benefit from newly introduced hard drives by educating them about SSD technology or adopting a website design that helps consumers discover the SSD subcategory. Put another way, removing the consideration barrier is key: once consumers consider the SSD category, they will find something they like despite search frictions. I also explore how consumer surplus changes when I reduce consumers' ability to search in a directed way, and I find that the surplus from SSDs reduces substantially. When consumers are forced to search in a random order, they can only discover and learn about SSDs by chance, which reduces the surplus from SSDs to \$1.2. This result can be interpreted to mean that currently available search tools help consumers discover and learn about new hard drives, thus increasing consumers' benefits of new product introductions.

One may wonder whether the proposed model is a reasonable way to account for information frictions. I believe that the two frictions I model are both realistic and grounded in empirical evidence. The idea that consumers limit their search to a specific category seems natural: one can imagine a person who is looking for a bottle of wine from a specific region (e.g., French Bordeaux), trying to book a hotel room in a specific neighborhood of a vacation town, or searching for a restaurant of a certain cuisine type (e.g., Italian or Japanese). The first model of such category consideration is proposed by Ching et al. (2009). Manzini and Mariotti (2012) study the theoretical revealed preference properties of such a model. Additionally, consider-then-choose models perform remarkably well at rationalizing anomalies in choice experiments (Manzini and Mariotti, 2010), fitting grocery purchase data (Ching et al., 2009), and explaining weak responses to price promotions (Seiler, 2013). Similarly, search frictions have been well-documented in a large variety of markets (Honka et al., 2019). Given that both frictions seem to be inherent features of consumer behavior, it makes sense to account for both category consideration and costly search when estimating demand for new goods.

This paper contributes to two key strands of literature. The first shares my primary goal of estimating consumer surplus from new goods. Ever since the seminal paper of Hicks (1940), this literature studied the value of new goods in a variety of markets, including those for computers (Bresnahan, 1986), breakfast cereals (Hausman, 1996), automobiles (Petrin, 2002), and books (Brynjolfsson et al., 2003). I show that to obtain plausible surplus estimates, the researcher needs to account for information frictions. The perfect information model misestimates surplus for two reasons: it fails to correctly predict what consumers would have chosen if the new product were not introduced, and it uses an incorrect surplus function that does not account for frictions. Since it is difficult to anticipate the sign of the resulting bias, researchers need to empirically account for information frictions in each application. The model I develop helps researchers account for such frictions, and my application of hard drives illustrates how that can be done in practice. Accounting for frictions also reduces the strong dependence of consumers' choices on unobserved product-specific shocks, thus partly removing the property considered undesirable in the discrete choice literature (Petrin, 2002; Ackerberg and Rysman, 2005; Berry and Pakes, 2007).

The second related strand of literature is on modeling consideration and search. Several authors develop empirical models of consumer search and estimate them using online browsing data (De Los Santos et al., 2012; Honka, 2014; Ursu, 2018; Moraga-González et al., 2021; Donnelly et al., 2021). My contribution is in extending these models with a category consideration stage, developing a Bayesian estimation method, and using estimation results to study consumer surplus from new products. To the best of my knowledge, the only other paper that jointly models consideration and search is Honka et al. (2017). Their work differs from mine in both the research question and methodology. While they focus on studying how advertising affects consumer choices, I focus on estimating consumer surplus from new products. They also pursue a different identification argument. To identify whether consumers are unaware of certain brands, which is analogous to category consideration in my model, they directly ask consumers in a survey which brands they are aware of and can recall. By contrast, I examine what consumers do when not shopping for hard drives, and I use this information to construct novel shifters of preferences, category consideration, and search costs. I also rely on an exclusion restriction that removes price from the consideration stage, similar to the identification argument in the consideration model of Ching et al. (2009) and Ching et al. (2014). It is worth noting that in addition to category consideration, Ching et al. (2014) model consumer learning about product quality, a channel closely related to search models. However, they do not use their model to measure the welfare benefits of new product launches. In this sense, my paper can be viewed as a complement to their research.

2 Market and Data

2.1 Hard drive market

The hard drive market provides an ideal setting to study the value of new products under information frictions. Given product turnover and large product assortments, consumers face a non-trivial search problem. Since repeat purchases are rare in this market, consumers effectively encounter a new search problem each time they return to the market to buy another hard drive. Additionally, this market constantly evolves through technological innovations that allow manufacturers to produce faster and more compact hard drives with greater storage capacities (Christensen, 1993; Igami, 2017). One recent major innovation came from introducing the conceptually new technology of solid state drives (SSDs). The traditional HDDs consist of circular platters that rotate at high speed, allowing the read-write heads to access information in different platter segments. By contrast, the newly introduced SSDs have no moving mechanical components and are based on NAND flash memory. This ensures SSDs are significantly faster than HDDs, more durable, and more resistant to physical shock.

Despite the promising new technology, consumers have been slow to adopt SSDs. While online retailers started offering mass-market SSDs in 2010-2011, by 2016 only 22% of hard drive buyers purchased an SSD.[3] One potential explanation is related to consumer preferences. Since SSDs are considerably more expensive, consumers may not be willing to pay for the additional speed that these hard drives offer. Consumers may also prefer to buy a high-capacity HDD rather than a low-capacity SSD if they require additional storage space. Another potential reason for this limited adoption is that consumers are imperfectly informed about SSDs. They may not consider SSDs either because they are unaware of this new technology, or because they simply misunderstand it. In fact, a quick online search reveals that PC users often get confused about the SSD technology, claiming that solid state drives are incompatible with many modern computers or that SSDs make it impossible to recover data in case of a hard drive failure. While neither of these claims is true, such misconceptions could prevent consumers from considering SSDs.

2.2 Click-stream data

I use click-stream data from the Comscore Web Behavior Panel. The data include the complete browsing histories of 81,418 U.S. Internet users in 2016. These users were chosen at random by Comscore from the sample of 2.5 million U.S. households. Comscore users install software meters on their computers and give Comscore permission to track all of their Internet activity. The dataset therefore includes the complete browsing history of each user with URL addresses, the history of purchases on major e-commerce websites, and users' demographic variables including age, income, and household size.[4] Importantly, I observe the URL addresses of all visited pages, which makes

[3] All numbers come from the Comscore dataset I describe in the next section.

[4] Comscore data are depersonalized and anonymized in a privacy-compliant manner.

Attribute	All drives ($N = 1,774$)		HDDs ($N = 1,346$)		SSDs ($N = 428$)	
	Mean	S.E.	Mean	S.E.	Mean	S.E.
Storage TB	1.28	1.63	1.51	1.78	0.57	0.65
Speed MB/s	180.1	161.6	94.6	35.8	406.7	145.5
Internal Drive	0.670	0.470	0.606	0.489	0.869	0.338
Brand: Seagate	0.184	0.387	0.230	0.421	0.040	0.196
Brand: WD	0.167	0.373	0.215	0.411	0.019	0.136
Brand: Toshiba	0.067	0.249	0.079	0.269	0.028	0.165
Brand: Samsung	0.058	0.234	0.022	0.145	0.173	0.379
Brand: SanDisk	0.024	0.154	0.001	0.027	0.098	0.298
Brand: Crucial	0.016	0.125	0.000	0.000	0.065	0.248

Table 1: **Attributes of hard drives that were offered on amazon.com in 2016.**

these data more detailed than most Comscore datasets used in prior work.[5] This highly granular data enable me to recover the exact list of hard drives each user examined and the order in which the user examined them.

To construct the main sample, I identify all users who shopped for hard drives on Amazon.com in 2016. Since more than 75% of hard drive purchases in my data were made on Amazon, this sample captures the majority of searches and purchases in this market. I first recover the complete list of 1,774 hard drives that were available on Amazon in 2016 and identify all users who "searched" at least one hard drive by opening its product page. I define a "search session" as all searches made within a week leading to a purchase; and for users without purchases, I define a search session by finding a week with the largest number of searches. Because users rarely make repeat purchases and conduct all their searches within narrow time intervals, these definitions leave me with a dataset where each user has exactly one search session. As a result, I obtain a sample of 2,422 users who searched a total of 4,366 unique hard drives. Only 222 of these users purchased a hard drive from Amazon, implying a conversion rate of 9.2%. Throughout this section, I conduct the analysis using the complete sample. In Section 4, when estimating the structural model, I simplify estimation by focusing on the 100 most frequently purchased hard drives. For additional details of data construction, see Appendix A.

Users search relatively little. Despite the large assortment available on Amazon, the average user only searches 1.8 hard drives. This observation suggests that either users face substantial information frictions, or they have strong preferences for certain hard drive attributes that make them reluctant to search beyond a few specific options. Importantly, around 70% of users do not search any SSDs during their search session, and SSDs represent only 21.1% of all purchases and 24.8% of searches in the main dataset. In the following sections, I study to what extent this reluctance to search SSDs is driven by preferences, category consideration, or costly search.

[5] De Los Santos et al. (2012) and De los Santos (2018) use another version of the Comscore dataset in which they only observe which online stores users visit but not which product pages they visit in each store.

	Price Mean	Price SD	Days on Promotion	Promotion Depth
All drives ($N = 1,774$)	\$140.60	\$12.40	77	-9.9%
HDDs ($N = 1,346$)	\$115.30	\$7.70	72	-10.3%
SSDs ($N = 428$)	\$220.30	\$27.30	93	-9.0%

Table 2: **Depth and frequency of price promotions for SSDs and HDDs.** The table shows the average prices of hard drives, the average standard deviation of prices across days, the number of days they were sold with at a discount (i.e., a price below median), and the average discount depth.

2.3 Attributes and prices

I also collect hard drive attributes from Amazon product pages and daily prices of hard drives from a third-party price tracking website (see Appendix A for details). My general strategy, consistent with the identification strategy in Section 3.4, was to collect all attributes that are visible to users on the product category page. I therefore collected information about each hard drive's price, brand, memory type (HDD or SSD), disk type (internal or portable), storage capacity in terabytes, and hard drive speed in megabytes per second. Table 1 summarizes these attributes. SSDs are, on average, faster and more compact than HDDs but offer lower storage capacity and more often require internal installation. While Seagate and Western Digital offer the largest assortments of HDDs, the category of SSDs is dominated by Samsung and SanDisk. Note, however, that the attributes in the two subcategories substantially overlap. Most brands offer both hard drive types, and the assortment includes plenty of fast HDDs as well as SSDs with high storage capacity. This attribute overlap will prove crucial for identifying category consideration (see Section 3.4).

Table 2 additionally shows that SSDs are almost twice as expensive as HDDs, but both hard drive types go through frequent promotion periods. An average SSD is on promotion 93 of 366 days (7-8 days each month), with an average discount of 9%. HDDs show comparable price variation. As I explain in Section 3.4, these temporary promotions help me identify category consideration.

Users focus their search on hard drives with similar attributes. Table 9 in the Appendix illustrates that users consistently search hard drives of the same brand and rarely switch to searching a different brand within the same session. The same observation can be made about other attributes. For instance, 90.4% of HDD searches are immediately followed by another HDD search, and 62.4% of SSD searches are immediately followed by another SSD search. While this search persistence might reflect that each user only considers one category, it might also reflect that users have heterogeneous preferences for attributes that distinguish HDDs from SSDs (e.g., storage capacity and speed). As I explain in Section 3.4, I leverage this observed search persistence to identify taste heterogeneity.

2.4 User-level variables

The Comscore dataset gives me a unique ability to study what other websites and pages users browsed outside the category of hard drives. Using these additional browsing data, I construct shifters of preferences, search costs, and category consideration. I briefly describe these shifters here, but interested readers may consult Appendix A for details.

Taste shifters. I construct several variables capturing what other products users searched and purchased apart from hard drives. First, because brand preferences may translate across product categories, brands users searched and purchased in the past might help me predict what brands they will choose when shopping for hard drives. To this end, I collect information on whether users searched or purchased other products from the brands represented in the hard drive category. Second, I identify users who purchased a desktop computer, laptop, video camera, or video game console before shopping for hard drives. Since these purchases might indicate an increased need for storage space, they might predict whether a user will search hard drives with high storage capacity. Finally, I also identify users of file-sharing websites (e.g., torrents) who might need additional storage space for the files they download, as well as the users of cloud storage services (e.g., Google Drive) who might not need much storage space. Using these taste shifters helps me estimate preference heterogeneity, which is especially difficult to do in my application where I do not observe repeat purchases. One could think about these shifters as carrying similar information as panel data, although my estimation uses it in a simplified way, without explicitly modeling multi-category shopping).

Consideration shifters. I also construct several variables measuring users' expertise in computer hardware. To identify computer enthusiasts, I locate users who visit specialized websites about PC hardware as well as gamers who might naturally have more computer expertise. To this end, I identify users who visited at least one website related to PC hardware, video games, or e-sports (see Appendix A for details). These users, whom I refer to as *tech-savvy* users, are more knowledgeable about technology, so they are more likely to be aware of the SSD category and might better understand the benefits offered by SSDs. Consistent with this idea, Figure 1 (left panel) shows that both hardware enthusiasts and gamers are more likely to search at least one SSD during their search session than other users. Table 6 in the Appendix provides further evidence of this effect by regressing key outcome variables on the tech-savvy indicator (i.e., indicator that a user visited PC hardware or gaming websites at least once) as well as on search cost shifters described below. The results show that tech-savvy users are about 70% more likely to search at least one SSD during their search session, suggesting that they are indeed more likely to consider SSDs. While tech-savvy users search 30% more options, most additional searches are SSDs and only a few are HDDs. Therefore, tech-savvy users search more, which mostly reflects that they examine more SSDs. One might worry that tech-savvy users have substantially different preferences, thus violating the exclusion restriction. To examine this concern, I test whether tech-savvy users are more likely to purchase expensive hard drives and hard drives with high storage capacity within the set of hard drives they search (see Table 6 in Appendix A.3). I cannot reject the null hypothesis that tech-savvy and non-tech-savvy users make the same purchase decisions conditional on search; therefore, I do not find any evidence that these users have substantially different preferences.

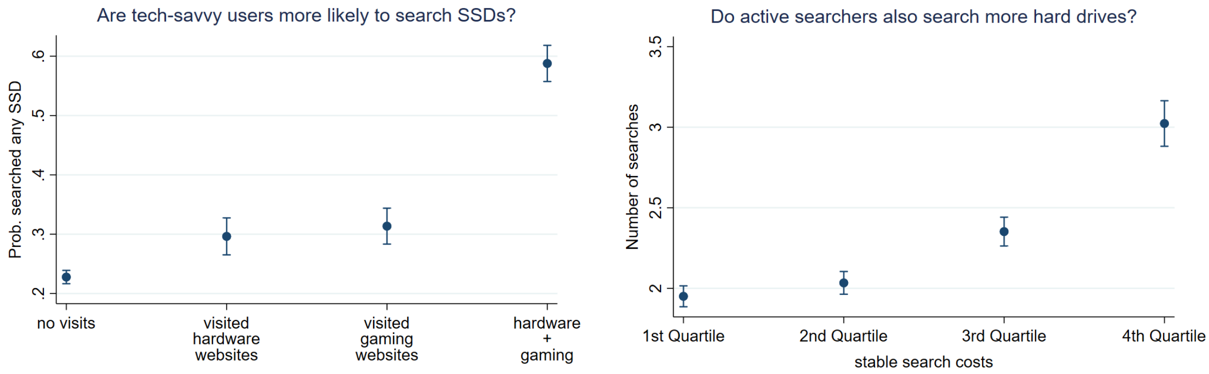

Figure 1: **Consideration and search cost shifters correlate with users' search decisions.** Both sets of shifters are introduced in Section 2.4. Appendix A.4 explains in detail how I identified visits to specialized hardware websites and gaming websites. Similarly, Appendix A.3 describes how I constructed the stable search propensity using individual-level search data from other product categories.

Search cost shifters. To construct search cost shifters, I study how much users search in other product categories on Amazon. While search costs may vary by context, product category, or even time of day, the search costs of individual users might be somewhat stable across categories. This stability might reflect the user's opportunity cost of time or their general ability to process information quickly and efficiently. I isolate this stable component by computing the average number of searches each user made in categories other than hard drives. In 2016, users visited about 3.5 million Amazon product pages of items other than hard drives. I classify all these 3.5 million products into categories by using category classifiers reported on Amazon product pages (see details in Appendix A). I then compute how many items each user searched in each category. Next, I regress the number of searches on user fixed effects and category fixed effects to partial out the impact of category-specific search costs. In what follows, I utilize the estimated user fixed effects to measure the so-called *stable search propensity* of this user. Figure 1 (right panel) illustrates that higher values of this search cost shifter are associated with more active search in the hard drive category on Amazon. Table 6 in the Appendix further shows that the preferences of these "active searchers" do not seem to systematically differ from those of other users.

I also attempted to measure how much time users spend on other online activities that compete for their limited attention (e.g., social media, emails and conference calls, news reading, etc.). I found that such time-use variables do not always correlate with search behavior in intuitive ways (see Appendix A.4 and Table 6 for details). Many of them also correlate with conditional purchase probabilities, suggesting they cannot be excluded from the preference equation. The only time use variable I include as a search cost shifter is the *daily total time online*, which is not correlated with conditional purchase probabilities. Users who spend more time online likely have a lower opportunity cost of time; therefore, they might have lower search costs.

3 Empirical Choice Model

The user starts the session by opening the hard drive category page, which presents available hard drives and their basic attributes (top panel in Figure 2). The user can then narrow down the list of options to specific categories (e.g. SSDs). Since I do not observe which categories users consider, the model treats category consideration as an unobserved process. The user then "searches" selected hard drives within the considered categories by opening product pages and reading detailed product descriptions (bottom panel in Figure 2). I capture this process using a sequential search model. After finishing search, the user buys one of the searched hard drives or leaves the website without buying.

3.1 Utility model

Consider a user who chooses from J available hard drives. The user i demands exactly one hard drive and derives the following indirect utility from buying a hard drive j:

$$u_{ij} = x_j'\beta_i - \alpha_i p_{j,t(i)} + \eta_{ij} + \varepsilon_{ij}, \tag{1}$$

where $p_{j,t(i)}$ is the price of hard drive j during the week of the visit $t(i)$, x_j is a vector of hard drive attributes that includes a constant, and α_i and β_i are the user's price sensitivity and tastes. Although Amazon does not personalize prices, I index price with $t(i)$ because different users visit the hard drive category on different weeks, thus getting exposed to different prices. I use this cross-sectional variation in prices to identify consideration separately from tastes, following the identification arguments outlined in Section 3.4. The specification in (1) assumes that hard drive prices do not change within the focal week $t(i)$, implying that within-week price changes are not used in estimation.

The vector x_j captures all time-invariant attributes that users observe in the category list (see the top panel of Figure 2). In my application, x_j includes the brand, type (internal vs portable), speed in megabytes per second, and storage capacity in terabytes. Additionally, Appendix E shows that model estimates are robust to the inclusion of hard drive quality and reliability in the utility function. I assume that the indirect utility in (1) is a linear function of attributes x_j. Relaxing this linearity assumption would be difficult because in a model with an arbitrarily flexible utility function, it would be hard to identify consideration parameters separately from preferences.[6] To ensure that preferences are identified, I also assume that both SSDs and HDDs are located in the same space of characteristics x_j. That is, I assume that prior to searching, users do not observe any other attributes not included in x_j that might discourage them from examining HDDs or SSDs. The terms η_{ij} and ε_{ij} in (1) are i.i.d. mean zero stochastic terms, normally distributed with variances σ_η^2 and σ_ε^2. In estimation, I normalize σ_ε^2 to fix the utility scale, and I estimate σ_η^2 together with

[6] An extreme example is a fully nonparametric model in which preferences are defined by $N \times J$ intercepts, one for each user-product pair. Such a model could perfectly rationalize any observed search behavior without the consideration stage, thus overfitting the data and making it impossible to test for limited consideration.

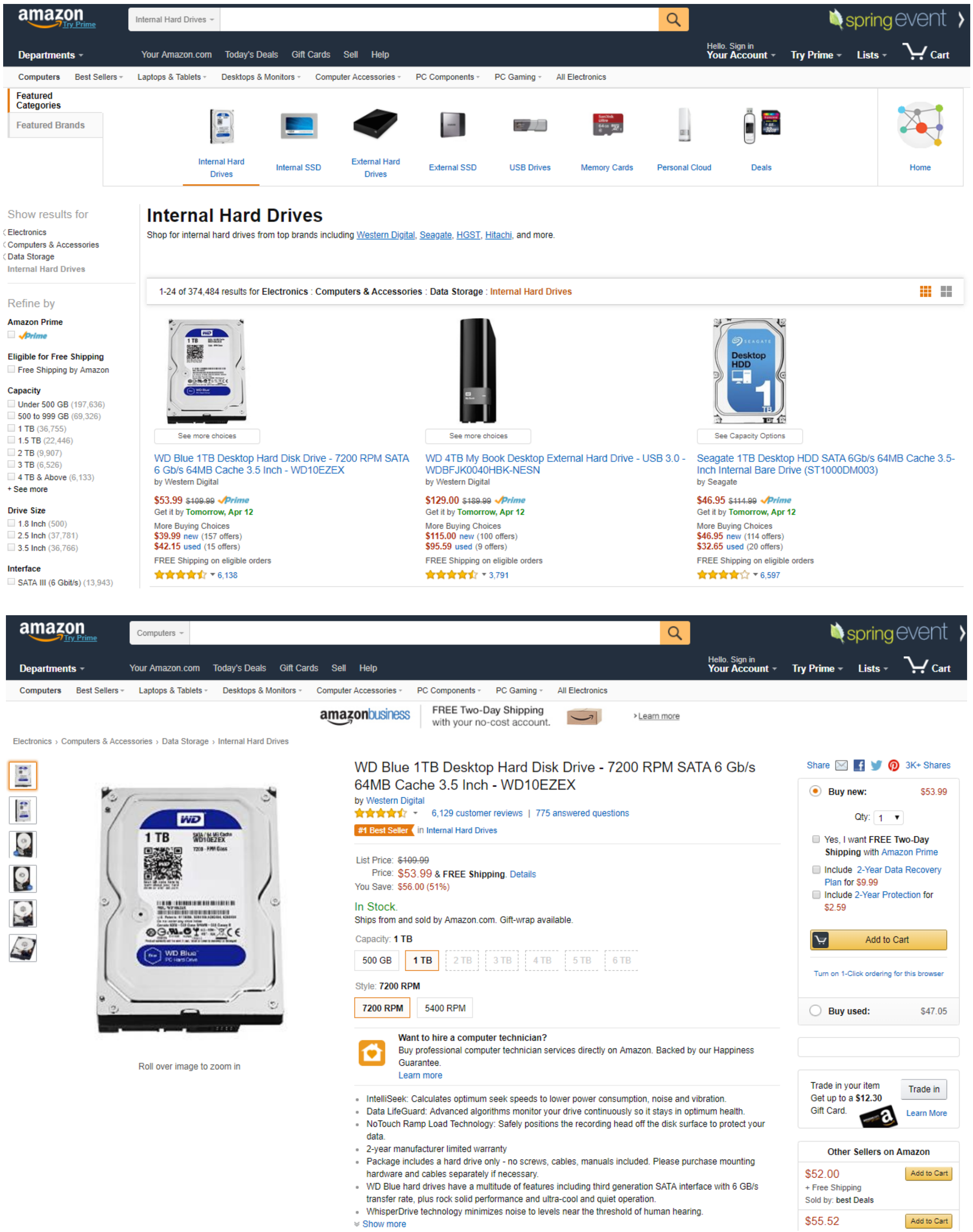

Figure 2: **An example of a category page (top panel) and a product page (bottom panel) as they appeared on amazon.com in 2016.** Source: The Wayback Machine (archive.org/web).

other parameters.[7] These stochastic terms affect search decisions in different ways, as explained in Section 3.2. Finally, the user can choose the outside option which yields utility $u_{i0} = \bar{u}_0 + \varepsilon_{i0}$ with $\varepsilon_{i0} \sim N(0, \sigma_\varepsilon^2)$. I normalize the mean utility of the outside option to zero for identification, so that $\bar{u}_0 = 0$.

The terms α_i and β_i capture user i's tastes and form this user's preference profile, $\theta_i = (\alpha_i, \beta_i')'$. To simplify estimation, I assume that each element k of this profile, θ_i^k, is independent from others and follows a univariate normal distribution such that $\theta_i^k \sim N(\pi_k' w_i^p, \sigma_k^2)$, where w_i^p is a vector of taste shifters. In principle, one can specify a more flexible distribution of types θ_i, but estimating such a flexible structure would likely require more data than I have. Although this way of modeling preference heterogeneity is standard in the literature, I extend user characteristics w_i^p to include novel demand shifters that reflect consumer behavior in other product categories. In my context, taste shifters w_i^p include variables capturing brand choices of user i in other Amazon categories, this user's recent purchases of other electronics products on Amazon, and the usage of file-sharing websites and cloud-storage solutions (see Section 2.4).

3.2 Information model

Category consideration stage. To model category consideration, I assume that the user i considers a hard drive j if and only if its *consideration index,* q_{ij}, is positive:

$$q_{ij} = d_j^{HDD} \gamma_i^{HDD} + d_j^{SSD} \gamma_i^{SSD} + \mu_{ij} \tag{2}$$

where d_j^{HDD} and d_j^{SSD} are HDD and SSD indicators for hard drive j, γ_i^{HDD} is the propensity of user i to consider a given HDD, γ_i^{SSD} is the propensity of this user to consider a given SSD, and the term $\mu_{ij} \sim N(0, \sigma_\mu^2)$ is the i.i.d. stochastic shock. All hard drives that have positive consideration indices q_{ij} constitute this user's consideration set, $C_i \subseteq J$. I model the heterogeneity in consideration propensities γ_i^{HDD} and γ_i^{SSD} by assuming that they follow univariate normal distributions such that $\gamma_i^m \sim N(\pi_m' w_i^c, \sigma_m^2)$ with $m = HDD, SSD$ where w_i^c are consideration shifters and σ_m^2 is the variance of unobserved heterogeneity.

Although consideration propensities γ_i differ across users, these propensities are stable within each hard drive type for a given user. This specification enables me to capture several realistic behaviors. A user might consider only hard drives of one specific type, for example, only HDDs. Such a user would always consider HDDs (e.g., $\gamma_i^{HDD} = +\infty$) and never consider SSDs (e.g., $\gamma_i^{SSD} = -\infty$), regardless of these hard drives' attributes. Similarly, a user might consider only SSDs but not HDDs ($\gamma_i^{SSD} = +\infty$ and $\gamma_i^{HDD} = -\infty$). The model also captures a continuum of cases between these two extremes. For instance, a user might be willing to consider both SSDs and HDDs but is a lot more likely to consider HDDs, which would correspond to the case of $\gamma_i^{HDD} \gg \gamma_i^{SSD}$. Apart from consideration propensities γ_i, the consideration sets are also affected

[7] Estimating the pre-search variance σ_η^2 enables me to express the estimated search costs in dollars. I would not be able to express search costs in dollars if both variances were normalized (Morozov et al., 2021; Yavorsky et al., 2021).

by idiosyncratic shocks μ_{ij} in equation (2). These shocks help me smoothen the consideration probabilities; therefore, they play a similar role to the pre-search and post-search utility shocks, η_{ij} and ε_{ij}. In this sense, the consideration model closely resembles the model of "soft" consideration sets in Goeree (2008).

As discussed in Section 2.4, tech-savvy users might be more informed about new technologies in this market and might therefore be more likely to consider SSDs. I capture this heterogeneity by including the "tech-savvy" indicator into the vector of consideration shifters w_i^c. The sign of the coefficients in π_{HDD} and π_{SSD} then determines whether tech-savvy users consider SSDs *instead* of HDDs or *in addition* to them. Importantly, equation (2) excludes prices, thus giving me an exclusion restriction necessary to identify consideration parameters (Section 3.4 further develops this point).

One can think of several reasons why users do not consider all hard drive categories. Users may be unaware of certain categories, misunderstand the category's benefits, find the category too complex to understand (e.g., in the case of the new and relatively unknown SSD technology), or they may eliminate certain categories to simplify the choice process. The consideration model in (2) is consistent with all these causes, and it remains agnostic about which of these drive users' choices in my application. In other words, this model uses "consideration" to capture all behaviors which make consumers limit the scope of their information search to a natural set of hard drives (e.g., the subcategory of HDDs). Modeling category consideration is in itself not a new idea, and it has been extensively explored in the theoretical literature (Manzini and Mariotti, 2012). Consideration models have also been shown to perform remarkably well at explaining consumer choices (Ching et al., 2009; Manzini and Mariotti, 2010; Seiler, 2013). Consistent with these results, I find that adding the consideration stage to the model substantially improves out-of-sample fit.

An advantage of using the "tech-savvy" indicator as a consideration shifter is that I can compare the behavior of users to that in the group of "expert" users who are more likely to be informed about the available hard drive technologies. A potential limitation, however, is that I need to assume that the "tech-savvy" indicator, included in the vector of consideration shifters w_i^c, does not correlate with the user's preferences (α_i, β_i). While I provide an indirect test that supports this assumption in Section 2.4, there does not seem to be a natural way to test this assumption directly with my data. Developing such a direct test is an important venue for future research.

The proposed model views category consideration as a passive process unrelated to preferences. In principle, one could endogenize consideration by assuming that users choose which categories to consider while anticipating future search. Such a model would be substantially more complex to specify and solve, as users would need to form expectations over what hard drives they will discover and search in each category. It would also be difficult to estimate such a model with the data I have, as I do not observe the actual consideration decisions. I abstract away from these additional complexities and assume that consideration only depends on exogenous consideration shifters.[8]

[8] Honka et al. (2017) similarly interpret consideration (which they term "awareness") as a passive occurrence unrelated to preferences, whereas they view search as an active learning process through which consumers gather information about available options.

Search stage. After choosing the consideration set $C_i \subseteq J$, the user proceeds to the search stage. I assume the user knows the attributes x_j, prices $p_{j,t(i)}$, and realized shocks η_{ij} for all hard drives in the consideration set C_i. That is, she knows the first part of the utility in (1), $\delta_{ij} = x_j'\beta_i - \alpha_i p_{j,t(i)} + \eta_{ij}$, which I term *pre-search utility.* One can think of a user who opens the category page of SSDs and learns the basic attributes of hard drives in that category. The shocks η_{ij} in this case capture user's preferences for any attributes that are difficult to quantify but that make certain options relatively more attractive (e.g., product photos). The user then proceeds to the search stage in which she searches hard drives one by one, revealing their ε_{ij} values that capture additional information displayed on product pages. By learning the values of ε_{ij}, the user effectively learns the realized utility $u_{ij} = \delta_{ij} + \varepsilon_{ij}$ of hard drive j, which I term the *post-search utility.*

I model search using the standard sequential search model of Weitzman (1979). Endowed with rational expectations about the distribution of ε_{ij}, the user examines hard drives one by one, revealing their ε_{ij} values and deciding at each step whether to continue searching. The first search is free, but the user incurs a fixed cost $c_i \geq 0$ for each subsequently searched hard drive. At any point during this process, the user can either purchase one of the searched options with realized utility u_{ij} or choose the outside option. The utility of the outside option, u_{i0}, is known to the user before search. Weitzman (1979) derives the optimal search behavior in this model. Define the reservation utility z_{ij} of hard drive j as a unique solution to the equation $\int_{z_{ij}}(u_{ij} - z_{ij})dF(u_{ij}|I_{ij}) = c_i$, where $I_{ij} = \{x_j, p_{j,t(i)}, \eta_{ij}\}$ summarizes what the user knows about the drive j before searching it, and $F(u_{ij}|I_i)$ is the distribution utility given this knowledge. This reservation utility z_{ij} captures the level of utility that makes the user indifferent between terminating search and searching the hard drive j. The user searches hard drives in the order of descending reservation utilities z_{ij} within her consideration set C_i. She continues searching as long as at least one unsearched product in C_i has a reservation utility above the utility in hand. Once the user finishes searching or exhausts all search opportunities, she chooses one of the searched options or the outside option, whichever yields higher utility.

I choose the sequential search model for several reasons. Since the model endogenizes search order, it helps me infer users' preferences from the order in which they search hard drives. Doing so helps me estimate preferences more precisely than would be possible with only data on the identities of searched products. This model also unlocks certain key counterfactual questions. For example, in Section 4.4, I ask how much surplus would SSDs generate if users could not search in a directed manner. This question would be difficult to answer in other search models which do not explicitly capture directed search. Lastly, this model assumes that individual tastes α_i and β_i are stable throughout the entire search session, which helps me abstract from complex models of consumer learning (Bronnenberg et al., 2016; Hodgson and Lewis, 2020).

For estimation, it helps to re-parametrize the model in the following way. I decompose the reservation utilities as $z_{ij} = \delta_{ij} + \xi(c_i)$ where $\xi(c_i)$ is a function that monotonically decreases in c_i and that only depends on the distribution of ε_{ij} (see the proof in Appendix B.1). Since the distribution of ε_{ij} is fixed by the scale normalization $\sigma_\varepsilon^2 = 1$, modeling the heterogeneity in search costs c_i is equivalent to modeling the heterogeneity in $\xi_i = \xi(c_i)$ which I term *search propensities*. Therefore, in practice I first estimate the distribution of ξ_i and then recover the implied distribution of search costs c_i after estimation. As I show below, this alternative parametrization generates inequalities that are linear in search propensities ξ_i, which simplifies the process of taking posterior draws. Operationally, I assume that propensities ξ_i differ across users such that $\xi_i \sim N(\pi_s' w_i^s, \sigma_s^2)$ where w_i^s are search cost shifters and π_s are corresponding regression coefficients. In my application, search cost shifters w_i^s include the stable search propensity as well as the average daily time spent online (see Section 2.4).

3.3 Bayesian estimation

Suppose we observe N users and know which hard drives each of them searched and in which order, as well as what they purchased. The goal of estimation is to use these individual search and purchase data to recover the unknown parameters of the model. It helps to think about this estimation problem as recovering the distribution of user types λ_i which include the price coefficient α_i, tastes β_i, search propensities ξ_i, and consideration parameters γ_i^{SSD} and γ_i^{HHD}. Given the assumptions in Sections 3.1 and 3.2, each element λ_i^k of the user's type is independent from others and follows the normal distribution $\lambda_i^k \sim N(\pi_k' w_i^k, \sigma_k^2)$. The goal of estimation is then to recover the regression coefficients π_k, heterogeneity variances σ_k^2, and the variance of the pre-search shock σ_η^2 from the data.

I estimate these parameters using a Bayesian approach. To the best of my knowledge, mine is the first paper that estimates a structural search model using Bayesian methods.[9] What makes the estimation of search models challenging is that, in most cases, the likelihood function does not admit a closed-form solution and has to be approximated. Many authors deal with this issue by using simulated likelihood methods (Honka, 2014; Honka and Chintagunta, 2015; Ursu, 2018). But since the likelihood of observing a specific search sequence is minuscule, one needs an extremely large number of draws to precisely approximate the likelihood. And because the resulting likelihood function is not smooth, the researcher also needs to introduce artificial smoothing (e.g., via kernel-smoothed simulator) which might bias the estimates (Chung et al., 2018). These issues would be even more troublesome in my application, where large assortments and user heterogeneity make the likelihood function even more complex and difficult to approximate.

To address these challenges, I develop a Gibbs sampler that approximates the posterior distribution of parameters using MCMC simulation methods. I find this approach appealing for several reasons. First, the MCMC sampler replaces the task of maximizing the likelihood with an algorithm that repeatedly draws from a series of conditional posterior distributions, thus removing the need to approximate the likelihood function. I found that in practice this makes the MCMC method more

[9] Yang et al. (2015) use Bayesian methods to estimate a boundedly rational model in which consumers engage in costly information search to learn product attributes. They model myopic consumers who make each search decision as if this were their last opportunity to gather information. By contrast, I develop Bayesian estimation for a rational search model in which consumers follow a dynamically optimal search strategy (Weitzman, 1979).

numerically stable and more robust to poorly selected starting values than likelihood-based alternatives. Second, the Bayesian approach is efficient at handling rich user heterogeneity (Rossi et al., 2012), and it provides a natural way to quantify uncertainty in finite-sample estimates without relying on asymptotic approximations.

To introduce the MCMC sampler, I first summarize the restrictions that observed searches and purchases impose on the model's parameters. Suppose the user i searches K_i hard drives and then purchases some hard drive y_i. Without loss of generality, let $1, \ldots, K_i$ be the indices reflecting the order in which the hard drives are searched, and with some abuse of notation, let $S_i = \{1, \ldots, K_i\} \cup \{0\}$ denote the resulting search set.[10] In what follows, I assume the search set S_i always includes the outside option. The observed decisions are optimal if and only if the following inequalities hold:

$$
\begin{aligned}
& j \in C_i \iff q_{ij} \geq 0 && \forall j && \text{(consideration)} \\
& z_{i,1} \geq z_{i,2} \geq \cdots \geq z_{i,K_i} \geq z_{i,j} && \forall j \in C_i \setminus S_i && \text{(search order)} \\
& \max(u_{i,0}, u_{i,1}, \ldots, u_{i,m-1}) \leq z_{i,m} && \forall m \in \{1, \ldots, K_i\} && \text{(continuation)} \\
& \max(u_{i,0}, u_{i,1}, \ldots, u_{i,K_i}) \geq z_{i,j} && \forall j \in C_i \setminus S_i && \text{(stopping)} \\
& u_{i,y_i} \geq u_{i,j} && \forall j \in S_i && \text{(purchase)}
\end{aligned}
$$

One can interpret these inequalities as follows. The user only considers hard drives with positive values of the consideration indices q_{ij} (consideration). Additionally, the user searches hard drives in the order of descending reservation utilities z_{ij} (search order), keeps searching until reaching the hard drive K_i (continuation), and stops after searching the hard drive K_i because the current utility exceeds reservation utilities of all unsearched hard drives in the consideration set (stopping). Finally, the user then either buys a hard drive or chooses the outside option, thus selecting an option from the set S_i (purchase).

The MCMC sampler takes sequential draws from the posterior distribution of unknown parameters while respecting this system of inequalities. To construct a quick sampler, I use the data augmentation technique and treat utilities δ_{ij} and u_{ij}, consideration indices q_{ij}, and user types λ_i as additional parameters to be estimated. The sampler then imputes the values of these additional parameters together with the structural parameters of interest. As a whole, this approach can be viewed as extending the standard hierarchical probit model by incorporating two information frictions, category consideration and costly search (Rossi et al., 2012, p.75). In fact, my model nests the perfect information probit model with i.i.d. shocks ε_{ij} as a special case when users face zero search costs and consider all hard drives ($c_i = 0$ and $\gamma_i^{HDD} = \gamma_i^{SSD} = +\infty$ for all users i).

[10] A more correct but cumbersome notation would be to interpret the observed search sequence as a pair $\{S_i, \pi_i\}$ where $S_i \subseteq C_i$ is the set of searched hard drives and π_i is a one-to-one mapping from S_i to the set of natural numbers $\{1, \ldots, |S_i|\}$ capturing the order in which these hard drives are searched.

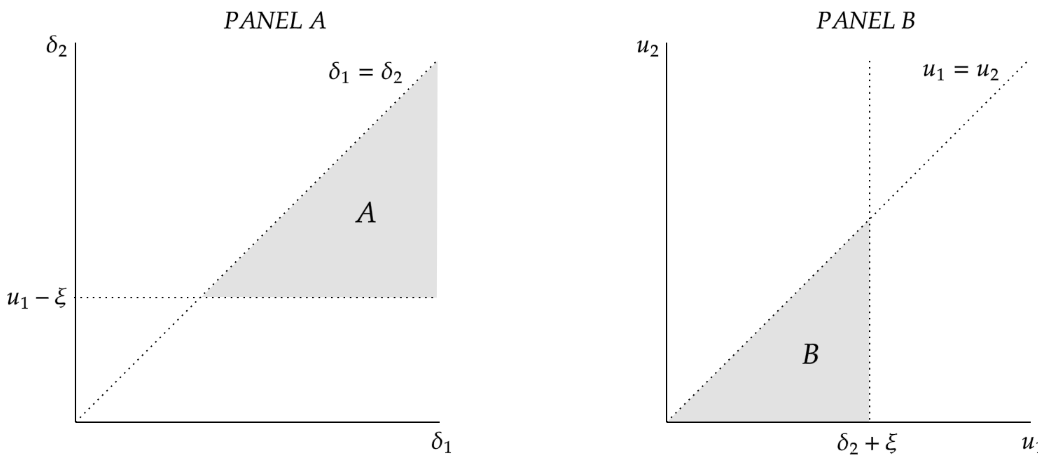

Figure 3: **Illustration of the proposed MCMC sampler.**

Implementation. One practical issue is how to generate draws of utilities (δ_{ij}, u_{ij}) for each user-product combination while satisfying a large number of non-linear inequalities. I deal with this computational challenge by using two tricks. First, decomposing the reservation utilities as $z_{ij} = \delta_{ij} + \xi_i$ makes inequalities linear in all parameters, which removes the need to deal with non-linear functions of search costs c_i. I recover the distribution of propensities ξ_i and infer the implied distribution of search costs c_i after estimation. Second, I simplify the system of inequalities, reducing it to a much simpler one. With this simplified system, drawing utilities δ_{ij} and u_{ij} and search propensities ξ_i becomes as basic as taking draws from truncated normal distributions. Since taking these draws is easy, I am able to impute unobserved utilities for all user-product combinations relatively quickly. This feature allows me to develop a practical MCMC sampler that scales well with large numbers of users and items. See Appendix C where I present the simplified inequalities, discuss the selection of priors, and derive all relevant posterior distributions.

I have tested the proposed MCMC sampler in several ways. I first used simulated search and purchase data to show that the MCMC sampler can successfully recover the true parameter values. To make these simulations informative, I made simulated data look similar to my actual sample in terms of the users' search and purchase behavior. While recovering the model's parameters is of some interest on its own, my final goal is to study welfare gains from new products, which are highly nonlinear functions of estimated parameters. For this reason, I also verified that the MCMC sampler can successfully recover the change in consumer surplus from new product introductions. See Appendix D for details.

Illustration. Suppose a user chooses between two items indexed by $j = 1, 2$ and assume there is no outside option. Assume that in the data, a user first searches item 1, then searches item 2, and then purchases item 1. This observed search sequence imposes the following constraints on item utilities:

$$\begin{aligned}
&\delta_1 \geq \delta_2 && \text{(search order)}\\
&u_1 \leq z_2 = \delta_2 + \xi && \text{(continuation)}\\
&u_1 \geq u_2 && \text{(purchase)}
\end{aligned}$$

The first inequality holds because searching item 1 before item 2 is only optimal when the reservation utilities are ordered as $z_1 \geq z_2$, which is equivalent to $\delta_1 \geq \delta_2$ given the flat search costs. The second inequality holds because the user decided to search item 2 after learning the utility of item 1, u_1. Finally, the third inequality holds because the user purchased item 1 which must have had higher utility than item 2.

Figure 3 visualizes these inequalities. Panel A shows all permissible values of pre-search utilities δ_1 and δ_2 conditional on utilities u_j and search propensity ξ. Similarly, panel B shows all permissible values of utilities u_1, and u_2 given the values of δ_j and ξ. To implement the proposed MCMC sampling, one would need to take draws of these utilities one by one, respecting the bounds depicted in these figures and conditioning each draw on all other unknown quantities (i.e, staying within the shaded areas A and B). For example, one can first draw δ_1 from the relevant normal distribution while truncating it with a condition $\delta_1 \geq \delta_2$, then draw δ_2 truncating it with $\delta_1 \geq \delta_2 \geq u_1 - \xi$, and so on. Once utilities δ_j and u_j are drawn, one can draw the search propensity ξ, truncating it with the condition $\xi \geq u_1 - \delta_2$. The actual MCMC sampling then boils down to generating posterior draws in this fashion for each user in the data.[11]

3.4 Identification

The main challenge is how to identify consideration parameters separately from search costs and preferences. My general strategy is to rely on two distinct sets of exclusion restrictions: one that builds on the user-level shifters w_i constructed in Section 2.4, and another one that excludes prices from the category consideration decisions. First, I utilize the user-level shifters w_i by isolating their effect on specific stages of the choice process. For instance, I include indicators for tech-savvy users as consideration shifters w_i^c, while at the same time assuming they do not impact the search process that follows consideration decisions. This approach is similar in spirit to the work by Bronnenberg et al. (2015) who compare the decisions of "expert" and "non-expert" consumers in order to isolate the impact of information on choices. Similarly, I include search cost shifters w_i^s that measure the total time spent online as well the average number of searches in other categories (i.e., "stable search propensity"). I assume that these variables correlate with search costs but do not directly influence the consideration decisions or the purchase decisions conditional on search. One such variable uses the extent of search in other categories as a shifter of search costs in the hard drive

[11] This sampling procedure is substantially different from the one used for estimating a full information probit model. In a full information model, one would be left with only one inequality $u_1 \geq u_2$ (item 1 is purchased) telling us that item 1 must have generated higher utility than item 2. By contrast, search data gives us a much richer system of inequalities, visualized in Figure 3.

category, which resembles the idea of Hausman instruments that use prices in other markets as shifters of prices in a focal market (Hausman et al., 1994).

Second, I exclude prices from the consideration stage, assuming that users do not react to price changes that affect hard drives they do not consider. This assumption resembles Ching et al. (2009) who also posit that consumers do not respond to promotions when they do not consider purchasing in a given category.[12] In contrast to their panel data approach, I use cross-sectional variation in prices across users. That is, although each user makes only one purchase, I observe similar users (e.g., tech-savvy users with the same demographics) visiting Amazon on different weeks and therefore seeing different prices. To use this cross-sectional variation, I assume that the week of the visit is uncorrelated with the user's preferences, i.e., tastes α_i and β_i are uncorrelated with visit time $t(i)$ in the utility model (1). I then identify category consideration from the extent to which users do not react to temporary price changes that affect specific hard drive types. For example, if exposing some of these users to lower SSD prices does not make them more likely to search SSDs, I interpret this as evidence of limited consideration.

Of course, differential responses to price changes might also be driven by heterogeneous price sensitivities α_i and tastes β_i. Since my model captures how such heterogeneity translates into choices, I use the model's structure to disentangle consideration from preferences. One may worry that, if SSD prices dropped dramatically, it would encourage many users to consider SSDs. From this perspective, it helps that virtually all variation in my data comes from temporary price discounts that reduce prices by only around 10% and are less likely to shift consideration (see Table 2). An additional worry is that Amazon encourages users to consider discounted products, thus directly affecting consideration. In Appendix E, I show that Amazon does indeed promote discounted hard drives by placing them in more salient positions, but the effect appears to be too small to influence my qualitative results.

Another challenge is how to separate the impact of search costs from that of preferences. Since search costs c_i (and therefore search propensities ξ_i are flat across items), users must search in the order of descending pre-search utilities δ_{ij} that depend only on preferences. Hard drives for which users have stronger preferences will then be searched earlier and more frequently. I can therefore identify the mean tastes $\bar{\theta}$ from the observed search order. The mean search propensity $\bar{\xi}$ can be then identified from the average number of searched hard drives.

To identify consumer heterogeneity, I mostly rely on modeling the heterogeneity in key parameters as a function of observed user characteristics w_i. These characteristics include user-level consideration and search cost shifters as well as other variables such preference shifters w_i^p (e.g., brand choices in other categories, searches and purchases of electronics products, etc.) and demographics (e.g., age, income, and household size). The observed heterogeneity coefficients π_k are then identified from the extent to which users with different characteristics w_i focus their search

[12] Honka et al. (2017) pursue a different strategy and identify consideration from survey data. Since my approach relies on a revealed preference argument rather than survey data, the two empirical strategies can be viewed as complementary. One could also imagine combining the two strategies, i.e. treating both survey data and limited price responses as noisy signals of consideration.

on different subcategories of hard drives (SSDs vs HDDs), search different numbers of options, and purchase hard drive with different characteristics conditional on search sets.

I identify unobserved preference heterogeneity σ_k^2 using search data. Ideally, I would have panel data where each user searches and purchases multiple times, but such data are unavailable due to the durable nature of hard drives. Fortunately, search data provides a "mini-panel" by describing different searches made by the same user, which helps identify unobserved preference heterogeneity.[13] If all users had the same preferences, they would search hard drives, on average, in the same order. By contrast, users with heterogeneous preferences would search different sets of hard drives, in the order that best matches their preferences α_i and β_i. We would then expect hard drives to have a lot more similar attributes x_j within search sets S_i of specific users than across search search sets of different users. This is indeed what I see in the data, and I document this pattern for different characteristics x_j in Appendix A.6. These search patterns help me identify the unobserved preference heterogeneity. In turn, the unobserved heterogeneity in search propensities ξ_i is identified from the variation in the number of searches across users, beyond what is predicted by the mean search propensity $\bar{\xi}$, mean tastes $\bar{\theta}$, and unobserved preference heterogeneity parameters σ_k^2.

Finally, I need to recover the variance of the pre-search shocks, σ_η^2. One can view the term η_{ij} in (1) as a structural error that determines how much hard drive prices and attributes affect search order decisions. As $\sigma_\eta^2 \to 0$, I obtain a "pure characteristic" model in which the order of search is fully driven by prices $p_{j,t(i)}$ and attributes x_j, whereas the purchase decisions conditional on search are stochastic due to the realized values of shocks ε_{ij}. By contrast, when σ_η^2 is large, both search order and conditional purchase decisions are stochastic and only weakly correlate with the hard drive attributes. Thus, I identify σ_η^2 from the extent to which the observed search order can be explained by the hard drive attributes and prices.

4 Estimation Results and Inference

4.1 Demand estimates from the full model

I first estimate the complete model with category consideration and costly search. Table 3 reports parameter estimates in the form of posterior means and standard deviations, whereas the last column in Table 4 shows the dollarized values of these estimates. Most coefficients are precisely estimated and have expected signs. Users prefer faster hard drives with higher storage capacity. An average user is willing to pay a premium of \$13.7 for buying a hard drive whose average read-write speed is higher by 100MB/s, and a premium of around \$82 for buying a hard drive with at least 1TB of storage capacity. Users are also willing to pay substantial premia for Seagate (\$26.1), Western Digital (\$17.9), and Samsung (\$66.8), in line with the fact that these three brands attract

[13] In this sense, the information in search data resembles second-choice data as in Berry et al. (2004). In second-choice data, the researcher knows which product a consumer would have purchased in the absence of the first choice. Similarly, in search data, the researcher knows which products a consumer searched but did not buy. If the purchased product became unavailable, the consumer would likely switch to buying another product from the same search set.

most searches and purchases. Therefore, an average user is willing to pay about \$65 extra for SSDs because these hard drives are faster (\$48.1 premium), mostly offered by Samsung (\$66.8 premium), but often have less than 1TB of storage space (negative \$51.6 premium). The implied SSD premium of \$65 makes sense given that the average price difference between SSDs in HDDs is \$58 in the estimation sample.

Users face substantial search frictions. Both estimated information frictions, category consideration and costly search, are large in magnitude. I estimate mean consideration parameters to be $\bar{\gamma}_{HDD} = -1.106$ and $\bar{\gamma}_{SSD} = -0.917$, implying that the average user considers a given HDD with a probability 13.9% and a given SSD with a probability 18.7%. Because HDDs in this sample are a lot more numerous than SSDs, these estimates imply that most hard drives that users consider are HDDs. The left panel of Figure 4 visualizes the predicted consideration sets for non-tech-savvy users. To generate this figure, I first randomly draw 1,000 consideration propensities γ_i^{HDD} and γ_i^{SSD} from their estimated distributions, and for each drawn pair of propensities I randomly draw 10,000 consideration sets $C_i \subseteq J$ from the consideration model in (2). Figure 4 visualizes the distribution of these randomly drawn consideration sets. We find that the average user considers only about 14-15 hard drives out of 100 available options, on average examining 11 HDDs and only 3 SSDs. In other words, the average user ignores about 90% of all SSDs available in this market.

I estimate the mean search propensity $\bar{\xi}$ to be 1.617. Inverting this estimate, I obtain that the implied mean search cost is \$1.5 with a standard error of \$0.18 (the formula for this inversion is derived in Appendix B.1). This cost is substantial given that the typical user only finds it worthwhile to search 1.7 hard drives out of 14-15 considered alternatives. As a result, the probability that a given user searches a specific SSD in this market is less than 2%. While I present a more detailed welfare analysis below, these estimates do suggest that frictions prevent users from fully internalizing the benefits of the introduction of SSDs.

I estimate substantial heterogeneity in tastes, search costs, and consideration parameters. Table 3 presents the estimates of unobserved heterogeneity for all included parameters as well as selected estimates of the observed heterogeneity. Most importantly, these heterogeneity estimates are consistent with the empirical strategy and exclusion restrictions outlined in Sections 2.4 and 3.4, implying that additional Comscore data indeed helps me to precisely estimate preference heterogeneity and separately identify different frictions. For example, I find that brand preferences generally translate across categories. For instance, a user who searched other Seagate products in that year is also twice more likely to search a Seagate hard drive. For the other two major brands, Western Digital and Samsung, I estimate this effect to be of the same magnitude.

I also find rich heterogeneity in consideration propensities and search costs. While the average search cost is \$1.5, it is only \$1.1 for users who spend 3-4 more hours online daily, and only \$0.6 for *active searchers* – i.e., users who search on average 2-3 more products in other categories on Amazon. It is reassuring to find that the search cost shifters proposed in Section 2.4 indeed strongly correlate with the estimated search costs. I also estimate search propensities to decrease with age and increase with income, suggesting that older and poorer consumers are most affected

	Posterior Mean	Posterior S.E.		Posterior Mean	Posterior S.E.
Price (100s of dollars)			*Seagate Brand*		
Mean	-1.505	(0.050)	Mean	0.393	(0.088)
Income \$40,000-75,000	0.314	(0.100)	Searched. Seagate Before	0.556	(0.148)
Income \$75,000-150,000	-0.329	(0.260)	Purch. Seagate Before	0.964	(0.631)
Income \$150,000+	-0.173	(0.294)	Unobs. Heterogeneity	0.705	(0.047)
Unobs. Heterogeneity	0.406	(0.048)			
			Western Digital Brand		
Constant			Mean	0.260	(0.066)
Mean	-3.890	(0.091)	Searched. WD Before	0.305	(0.147)
Unobs. Heterogeneity	0.532	(0.048)	Purch. WD Before	0.037	(1.091)
			Unobs. Heterogeneity	0.787	(0.075)
Speed (100s MB/s)					
Mean	0.207	(0.026)	*Samsung Brand*		
Unobs. Heterogeneity	0.144	(0.028)	Mean	1.005	(0.174)
			Searched. Samsung Before	0.145	(0.049)
Storage Capacity 1-2 TB			Purch. Samsung Before	-0.266	(0.698)
Mean	1.227	(0.081)	Unobs. Heterogeneity	1.918	(0.321)
Uses Torrents	0.276	(0.120)			
Uses Cloud Services	-0.021	(0.099)	*Search Propensity*		
Unobs. Heterogeneity	0.484	(0.066)	Mean	1.617	(0.041)
			Stable Search Propensity*	0.676	(0.218)
Storage Capacity 3+ TB			Time Online (hrs per day)	0.017	(0.008)
Mean	1.288	(0.082)	Unobs. Heterogeneity	0.469	(0.071)
Uses Torrents	0.171	(0.203)	Standard deviation σ_η	1.185	(0.026)
Uses Cloud Services	0.516	(0.120)			
Unobs. Heterogeneity	1.401	(0.124)	*Consideration HDDs*		
			Mean	-1.106	(0.027)
Internal Drive			Tech-savvy indicator**	-0.182	(0.071)
Mean	-0.130	(0.093)	Unobs. Heterogeneity	0.148	(0.010)
Unobs. Heterogeneity	0.354	(0.092)			
			Consideration SSDs		
			Mean	-0.917	(0.070)
			Tech-savvy indicator**	0.778	(0.115)
			Unobs. Heterogeneity	0.258	(0.024)

Table 3: **Parameter estimates from the model with search and consideration.** The table shows the estimated means and unobserved heterogeneity variances for all attributes x_j included in the model, and it additionally shows selected estimates of the observed heterogeneity. I obtain these estimates from the main sample of 1,852 users who made in total 2,872 searches. The means of all parameters correspond to the estimates for users with average the values of observed characteristics, $\bar{w}_i$. Prices are measured in hundreds of dollars, and hard drives' read-write speeds are measured in hundreds of megabytes per second. *Stable search propensity is a metric capturing the average number of searches of the same user in other product categories. (see Section 2.4 for details). ** Tech-savvy indicator is an indicator for users who read specialized PC hardware websites or gaming news online.

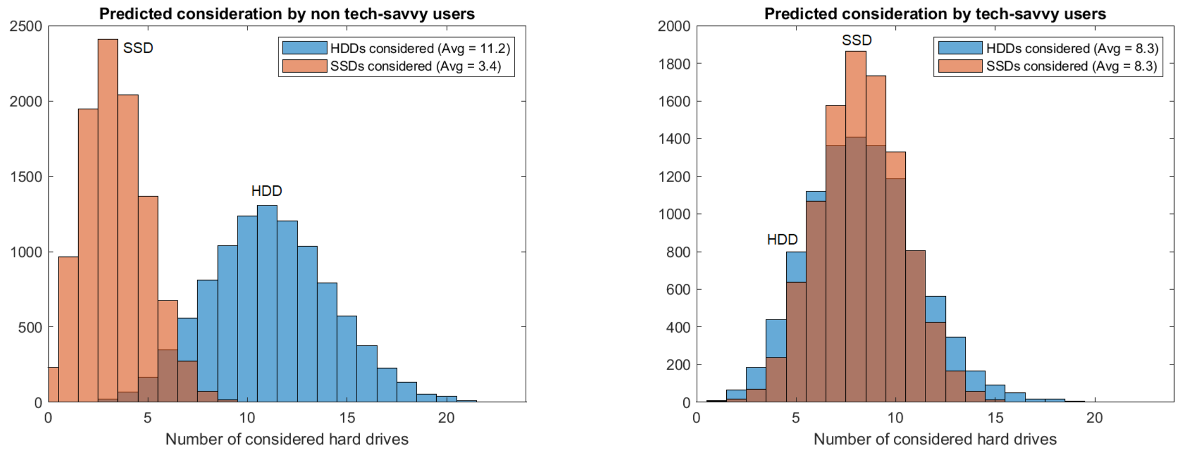

Figure 4: **Predicted consideration set sizes for tech-savvy and non-tech-savvy users.** I first randomly draw 1,000 consideration propensities γ_i^{HDD} and γ_i^{SSD} from their estimated distributions, and for each drawn pair of propensities I then randomly draw 10,000 consideration sets using the model (2). The figure on the left displays the predicted distribution of consideration set sizes for non-tech-savvy users, whereas the figure on the right shows the same distribution for tech-savvy users. Tech-savvy users are those users who, according to Comscore data, read specialized websites about PC hardware and gaming. The legend reports the average number of considered HDDs and SSDs for each user group.

by search frictions. Turning to consideration, I find that users identified as tech-savvy have different estimated consideration parameters. I estimate their probability of considering a given SSD to be 44.7%, which is 2.5 higher than that of non-tech-savvy users (18.7%). They also seem to consider the SSD subcategory *instead*, not *in addition* to HDDs, as their probability of considering a specific HDD (10.2%) is lower than that of other users (13.9%). The right panel in Figure 4 visualizes the predicted consideration sets for tech-savvy users and shows that these users' consideration sets tend to be equally split between SSDs and HDDs. Overall, I find that tech-savvy users are more likely to consider both subcategories during their search compared to other users, which makes them more likely to purchase an SSD.

4.2 Estimates under perfect information

The main question that motivated this paper is how researchers should estimate the value of new goods in markets with information frictions. Since I argue that it is critical to account for frictions, I need to compare my estimates with simpler models that the literature used for measuring the value of new goods. To this end, I estimate a perfect information model which is equivalent to the model in Section 3 but in which all users face zero search costs and consider both categories ($c_i = 0$ and $\gamma_i^{HDD} = \gamma_i^{SSD} = +\infty$ for all users i). This assumption essentially reduces the model to a perfect information probit model with random coefficients.[14] Since this model does not explain

[14]In the perfect information model, I drop the term η_{ij} from indirect utility and normalize the variance of ε_{ij} to one. Since both terms are additive, normally distributed, and known to the users, dropping one of them is without any loss of generality. As usual, all estimated coefficients are then identified relative to the normalized variance of

	Perfect Information Model			Model with Search & Consideration		
	Posterior Mean	Posterior S.E.	Dollarized Value	Posterior Mean	Posterior S.E.	Dollarized Value
Mean Preferences						
Price *(100s of dollars)*	-0.401	(0.063)	-	-1.505	(0.050)	-
Speed (100s of MB/s)	0.025	(0.009)	\$6.2	0.207	(0.026)	\$13.7
Internal Drive	-0.168	(0.034)	-\$41.9	-0.130	(0.093)	-\$8.6
Storage 1-2 TB	-0.129	(0.017)	-\$32.2	1.227	(0.081)	\$81.5
Storage 3+ TB	0.119	(0.049)	\$29.6	1.288	(0.082)	\$85.6
Seagate Brand	0.128	(0.030)	\$31.9	0.393	(0.088)	\$26.1
WD Brand	-0.026	(0.026)	-\$6.5	0.260	(0.066)	\$17.3
Samsung Brand	-0.140	(0.063)	-\$35.1	1.005	(0.174)	\$66.8
Unobserved Heterogeneity						
Price *(100s of dollars)*	0.118	(0.028)	-	0.406	(0.048)	-
Speed (100s of MB/s)	0.026	(0.008)	\$6.4	0.207	(0.026)	\$9.6
Internal Drive	0.121	(0.036)	\$30.1	0.354	(0.092)	\$23.5
Storage 1-2 TB	0.082	(0.032)	\$20.4	0.484	(0.066)	\$32.2
Storage 3+ TB	0.067	(0.022)	\$16.8	1.401	(0.124)	\$93.0
Seagate Brand	0.102	(0.055)	\$25.6	0.705	(0.047)	\$46.8
WD Brand	0.047	(0.026)	\$11.8	0.787	(0.075)	\$52.3
Samsung Brand	0.112	(0.091)	\$28.0	1.918	(0.321)	\$127.5

Table 4: **Comparison of Preference Estimates in Models with and without Information Frictions.** The table compares estimated tastes from the perfect information model which assumes zero search cost and full category consideration (columns 1-3) with those from the full model with category consideration and costly search. I obtain the estimates from the full model using the main sample of 1,852 users who made in total 2,872 searches. The estimated mean values of preferences α_i and β_i are reported for users with the average values of observed characteristics, $\bar{w}_i$. Prices are measured in hundreds of dollars.

why consumers search, I only use purchase data for estimation but preserve all 1,852 users in the sample.

Table 4 shows the results from this perfect information model. Compared to the search model, the perfect information model substantially underestimates users' preference for SSDs. This discrepancy is comprised of several biases. In the perfect information models, users are willing to pay only \$6.2 for buying a hard drive whose speed is 100MB/s higher, which is much lower than the \$13.7 premium in the full model with frictions. Instead of predicting a large and positive premium for Samsung, the most popular SSD brand, the perfect information model estimates a negative premium of \$35. When put together, the perfect information estimates imply that the average user prefers HDDs over SSDs (with a negative premium of \$2), which is at odds with the positive SSD premium of \$65 that I estimated in the main specification.

The two models yield different estimates for the following reasons. Because the perfect informa-

ε_{ij}.

tion model assumes users perfectly observe utilities of all products, it concludes that users do not buy SSDs because they find their attributes unattractive (e.g., high speed, Samsung brand). This incorrect inference leads to an underestimation of users' preferences for SSDs. In reality, however, the low market share of SSDs can be partly explained by information frictions. Some users do not consider SSDs altogether, while others consider but do search them due to high search costs. The full model in Section 4.1 correctly recognizes this and returns more plausible estimates of the users' preferences for SSD.

One may also ask whether modeling category consideration is even necessary for explaining search and purchase data. My estimates suggest that modeling consideration is indeed necessary. The full model shows good out-of-sample fit, which I discuss in Appendix F and illustrate in Table 14. In the same appendix, I also explore simpler variations of the same model. Notably, I explore a "search only" version of my model that assumes users consider all products but maintains the assumption of costly search. Such a model shows significantly worse out-of-sample fit, and it underpredicts consumer surplus from SSDs by around 30%. This bias arises because the model without category consideration incorrectly assumes all users consider SSDs; therefore, it erroneously interprets limited SSD searches as a signal that users value high storage space but do not value speed. For these reasons, I treat the full model as a preferred specification and use it for welfare analysis.

4.3 Welfare estimates

To measure how much users benefit from SSDs, I would ideally observe the hard drive market before and after the SSD introduction. However, my data only cover 2016 when SSDs were already available. Therefore, I need to predict users' choices and consumer surplus in a counterfactual scenario where SSDs are not introduced. To this end, I remove SSDs from Amazon's assortment, simulate users' search and purchase decisions from the estimated model, and calculate the change in consumer surplus.

I define consumer surplus as the ex-ante expected utility u_{ij} from the purchased hard drive, where the expectation is taken with respect consideration sets C_i, search sets S_i, purchase decisions y_i, and user types λ_i. In practice, computing consumer surplus takes a lot of time because I need to approximate a high-dimensional integral using simulations. To circumvent this issue, I use the result in Choi et al. (2018) to represent my model as a discrete choice model in which users choose an option with the highest *effective utility* defined as $v_{ij} = \min(u_{ij}, z_{ij})$. This representation removes the need to repeatedly solve the optimal search problem, thus simplifying welfare computations. I then compute the expected consumer surplus from the choice set J as follows:

$$CS(J) = E\left(\max_{j \in C_i}\{v_{ij}(\theta_i, c_i)\}\right) = \int \max_{j \in C_i}\{v_{ij}(\theta_i, c_i)\} dF(C_i|\lambda_i) dF(\lambda_i|\rho) \qquad (3)$$

where J denotes the set of available hard drives (e.g., $J = HHD$ in the counterfactual scenario without SSDs); $\max_{j \in C_i}\{v_{ij}(\theta_i, c_i)\}$ captures the highest effective utility among hard drives in the

Model	Taste Estimates	Unconditional		Cond. on purchase		Sources of ΔCS	
		ΔCS	% of price	ΔCS	% of price	Tastes	Shocks $\eta_{ij}, \varepsilon_{ij}$
Full Model	Full Model	\$3.20	3.3	\$25.10	26.0	57.7%	42.3%
		(\$0.60)		(\$4.60)			
Full Model	Perfect Info	\$0.70	0.7	\$5.50	5.6	58.2%	41.8%
		(\$0.10)		(\$0.80)			
Perfect Info	Perfect Info	\$1.20	1.3	\$9.60	10.0	40.9%	59.1%
		(\$0.20)		(\$1.80)			

Table 5: **Consumer surplus change from the introduction of SSDs.** The table reports the estimated change in the expected consumer surplus (ΔCS) after adding SSDs to the choice set of users, unconditional (columns 1-2) and conditional on a purchase (columns 3-4). The value of ΔCS is computed from the formula (10) as explained in Section 4.3. Consumer surplus after the introduction of SSDs corresponds to the consumer surplus predicted by the estimated model with frictions. By contrast, I compute the surplus before the introduction of SSDs by removing SSDs from the choice set and simulating users' decisions in this new environment (see Section 4.3 for details). I then compute the surplus change by taking a difference between two surplus estimates and dividing it by the average price coefficient. The last three columns further decompose this surplus change into the effects of different utility components (columns 4-6).

consideration set C_i; $F(C_i|\lambda_i)$ denotes the distribution over potential consideration sets $C_i \subseteq J$ given the type λ_i, and ρ is a vector of all estimated parameters. I approximate this expression using simulations (see Appendix G for details).

Table 5 shows the welfare estimates. Columns 1-2 report the unconditional surplus change, whereas columns 3-4 report the change in consumer surplus conditional on making a purchase. The estimates from the full model in row 1 imply that consumers derive substantial surplus from the introduction of SSDs, with the average surplus change of \$3.2, around 3.3% of the average hard drive price in the sample. The perfect information model, however, yields a substantially lower estimated surplus change of only \$1.2 (1.3% of the price), almost three times lower than that implied by the full model. Although both estimates seem small, note that the average user in the estimated model purchases only with 12% probability. Therefore, the surplus change conditional on purchasing a hard drive is as high as \$25.1 or 26% of the price (see columns 3-4).

Two main reasons explain why ignoring frictions leads to biased welfare estimates. First, when SSDs are introduced, users in the perfect information model immediately learn the shocks ε_{ij} (and therefore utilities u_{ij}) of all new products. At least one of these utility draws is likely to be more appealing than those of the HDDs. This is why the perfect information model predicts that users are more likely to buy SSDs than they would be under information frictions, thus overestimating consumer surplus. Consistent with this explanation, switching from the perfect information to the search model, while keeping preferences fixed, reduces the estimated surplus from \$1.2 to \$0.7 (rows 2-3 in Table 5).

That standard models overestimate gains from new products is a well-known result (Bajari et al., 2001; Petrin, 2002). One could address this by correcting the variance of taste shocks (Ackerberg and Rysman, 2005) or removing these shocks altogether (Berry and Pakes, 2007). My model offers an alternative solution. By explicitly capturing costly search and limited consideration, the model recognizes that users only observe the utilities u_{ij} of SSDs they actually decided to consider and search, which is typically a small subset of available SSDs. In fact, in the full model, the surplus from SSDs depends on shocks ε_{ij} a lot less than in the perfect information model (see the last column of Table 5). Thus, the full model is a more realistic model of consumer behavior in that the expected utility does not overly depend on the idiosyncratic shocks ε_{ij}.

The second reason for the observed bias is tied to the estimates of preferences. The perfect information model mistakenly attributes the low market share of SSDs to users' preferences, whereas in reality, this low share is partly explained by information frictions. As a result, the perfect information model substantially underestimates the users' preference for SSDs, thus also underestimating the surplus change. I illustrate this effect in rows 1 and 2 in Table 5, which show that switching to the "right" preference estimates from the full model increases the surplus change almost five times, from \$0.7 to \$3.2.

Overall, the surplus change is underestimated in the perfect information model because of implausible preference estimates. At the same time, this surplus is overestimated because the perfect information model overpredicts how many users discover an appealing SSD once this type of hard drives is introduced. In theory, the net effect of these two forces is ambiguous. Nevertheless, in my application, the model without information frictions underestimates the average consumer surplus from SSDs by a factor of three.

4.4 Information frictions and welfare

I have so far argued that to estimate surplus from SSDs, one needs to account for information frictions. I now turn to a related but different question of whether frictions affect users' ability to internalize the benefits from SSDs. Answering this question is informative because firms can potentially reduce these frictions via marketing activities (e.g., advertising), and Amazon can reduce them by adopting a website design that facilitates search and encourages product exploration. The following counterfactuals explore to what extent such changes could help users to more fully internalize the surplus from SSDs.

I present two counterfactuals: one that reduces the magnitude of information frictions from their estimated level, and the other one that makes frictions more extreme. First, I gradually remove the two frictions: category consideration and costly search. Figure 5 shows the resulting change in the estimated consumer surplus from SSDs. Scenario 4 in the middle indicates the surplus change in the estimated model, which is reproduced from Table 5. In scenarios 5 and 6, I increase the number of tech-savvy users from the actual 25% in the data to counterfactual values of 60% and then to 100%. To achieve this, I randomly sample non-tech-savvy users and draw their consideration parameters from the distribution of consideration propensities γ_i^{HDD} and γ_i^{SSD} I estimated for tech-savvy users in Table 3.[15] Such change could occur if hard drive manufacturers educated users about the benefits

[15] I avoid unrealistic counterfactuals that reduce the search cost to a very low value (e.g., zero) or increase the consideration probability to values close to 100%. Such counterfactuals generate unrealistic consumer behavior (e.g.,

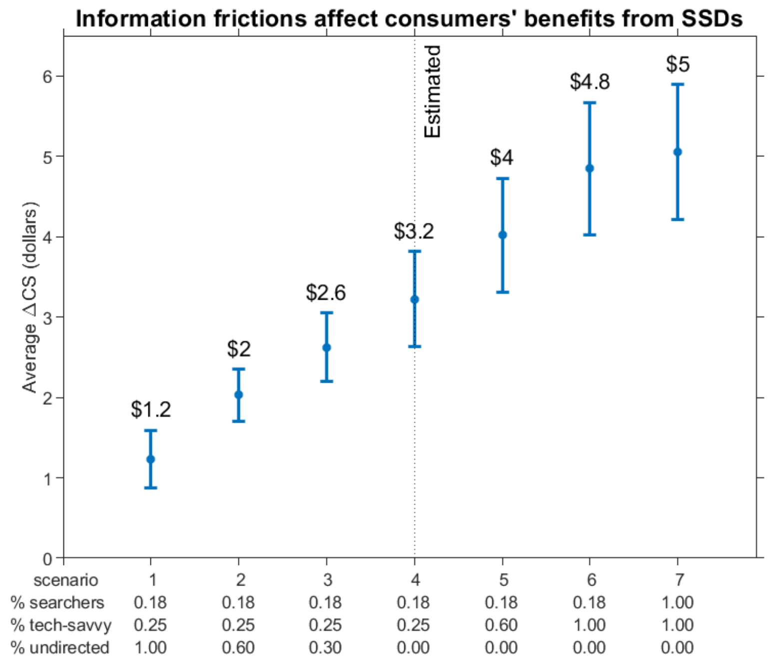

Figure 5: **Relationship between information frictions and consumer surplus from SSDs.** The graph shows the change in consumer surplus from adding SSDs to the assortment. Scenario 4 in the middle shows the surplus predicted from the estimated model. Scenarios 5-6 increase the number of tech-savvy users to 60% and 100%, and scenario 7 additionally reduces the search costs of all users to the level of search costs of "active searchers". Scenarios 1-3 add frictions by increasing the share of users who engage in undirected search.

of SSD technology, or if Amazon changed the category page layout to encourage the consideration of SSDs. Because making users tech-savvy also makes them consider a lot more SSDs, the average surplus from SSDs substantially increases to \$4 and \$4.8 (scenarios 5-6).

I then additionally reduce search costs of all users to the level of "active searchers" discussed in Section 4.1, thus reducing the average search cost from \$1.5 to \$0.6. While doing so increases the average surplus from SSDs to \$5, the change is relatively small compared to the impact of increasing consideration. This result is intuitive: in this model, users who need a high-speed hard drive can focus on searching SSDs with the most attractive attributes x_j (e.g., storage capacity and price), so higher search costs only affect how many SSDs they will search, not whether they search them at all. From this perspective, not considering the SSD category is more damaging to their welfare than a search cost increase.

Second, I also analyze scenarios in which I make it more difficult for users to do a directed search while keeping their search costs and consideration parameters fixed at the estimated levels. This scenario mimics the time before the introduction of search engines and online platforms, in which it was more difficult for users to focus their search on hard drives with desired attributes.

users searching 70-80 hard drives per session or buying with a probability close to one) and are not informative about the potential effects of marketing policies.

Operationally, I make some proportion of users search in an "undirected" way, assuming they do not observe attributes x_j and $p_{j,t(i)}$ before search but know the distribution $F(x_j, p_{i,t(i)})$ from which these attributes are drawn. In such a model users search in random order with respect to attributes x_j and $p_{i,t(i)}$, meaning they need to rely on luck or conduct extensive search in order to discover hard drives with desired attributes. Note that switching users from a "directed" to an "undirected" search may have complex effects on user welfare; for example, adding SSDs may actually decrease the surplus of users who only care only about HDDs, because they may now have to waste their searches on SSDs that do not match their preferences. While the optimal search rule is still the same as under directed search, the reservation utilities – call them $\tilde{z}_{ij}$ – are now determined by a different equation:

$$\int_{u_{ij} \geq \tilde{z}_{ij}} (u_{ij} - \tilde{z}_{ij}) dF(x_j, p_{i,t(i)}) dF(\varepsilon_{ij}) = c_i. \tag{4}$$

where the expectation over utilities u_{ij} is now with respect to the distribution of shocks ε_{ij}, attributes x_j, and prices $p_{i,t(i)}$. Assuming rational expectations, I estimate the distribution $F(x_j, p_{i,t(i)})$ using the empirical distribution of hard drive attributes and prices in the data. As before, I compute the consumer surplus using the simplified formula in (10), except that I now use the "undirected" reservation utilities $\tilde{z}_{ij}$ and compute the effective utilities as $\tilde{v}_{ij} = \min(u_{ij}, \tilde{z}_{ij})$.

Figure 5 shows the resulting surplus estimates (scenarios 1-3). I gradually increase the number of users doing undirected search by increasing their share from zero in the benchmark scenario first to 30% in scenario 3, to 60% in scenario 2, and finally to 100% in scenario 1. The results reveal that removing directed search substantially decreases the surplus from SSDs. The surplus drops from \$3.2 in the benchmark case, where all users search in a directed way, to \$2.6, then to \$2, and eventually decreases to \$1.2 in scenario 1 where all users search in an undirected way. Overall, these results suggest that the ability to do directed search, facilitated by the existing search tools, substantially increases the ability of users to benefit from SSDs.

When put together, the scenarios in Figure 5 range from a frictionless market to a market with severe frictions where users essentially search in random order. As we move toward cases with more frictions, SSDs generate less surplus, because it becomes increasingly difficult for users to discover SSDs that match their tastes. The generated surplus can be anywhere between \$1.2 and \$5 depending on the magnitude of frictions. Among other things, this observation suggests that any technology shifts that reduce frictions, such as internet penetration or new comparison platforms, may help users reap additional benefits from newly introduced hard drives. Whether such changes can, in turn, encourage manufacturers to develop and produce new hard drives is a fascinating question for future research.

5 Discussion and Conclusions

Many important economic questions hinge on the extent to which consumers benefit from new goods. By estimating benefits from new goods, researchers can provide insights about consumer surplus from product innovations (Hausman, 1996; Petrin, 2002), benefits from increasing product variety (Brynjolfsson et al., 2003), and gains from international trade (Broda and Weinstein, 2006). This paper provides an empirical framework for estimating consumer surplus from new goods under information frictions. Using the application of hard drives, I show that accounting for frictions is critical for obtaining plausible welfare estimates. I also argue that information frictions substantially diminish consumer gains from new goods. Broadly speaking, this result implies that frictions prevent consumers from fully internalizing the surplus from new product introductions.

My analysis has several limitations. Because of data restrictions, I only observe consumer behavior on Amazon but not in other online or offline stores. All results, therefore, apply to Amazon consumers and cannot be immediately extrapolated to the whole population of hard drive buyers. It would be helpful to construct more complete measures of the consumer search process, documenting how consumers search across stores and channels. Doing so might generate novel research questions such as whether the competition between stores creates an environment more conducive to new product discovery. Another limitation of my data is that I do not observe whether consumers limited their search to a subcategory of products, for example, by using a search query "solid state drives." Without this additional information, I can only infer category consideration indirectly from the observed searches and purchases. Next, it would also be interesting to study the extent to which consumers do not consider SSDs because this category features relatively unknown brands. While technically I could model consideration as a function of brand-specific variables (e.g., advertising), whether one can identify such a consideration model remains an open question. Finally, another limitation of my work is that in the model, Amazon plays a passive role in the consumer search process. In reality, Amazon may strategically design the website layout to promote certain hard drives. All analyses here should therefore be interpreted as conditional on the current website's layout.

For practical reasons, I abstracted away from several nuanced features of the search process that might characterize online shopping. While the model assumes that search costs do not vary across hard drives, in reality consumers may find it more difficult to gather information about certain hard drive types (SSDs). One could model such an environment by estimating product-specific search costs as in Ursu (2018). I leave this extension for future research.[16] I have also abstracted from the role of consumer learning and assumed that consumers' preferences remain stable during search. By contrast, recent empirical research shows that consumers often "zoom in" on a small set of similar products (Bronnenberg et al., 2016). While there have been some initial attempts

[16] To implement this extension, one would need to modify the Gibbs sampler in Appendix C by introducing product-specific search propensities ξ_{ij}. However, one would need to develop an empirical strategy that can identify product-specific search costs separately from consideration and preferences. An interested reader may consult Morozov et al. (2021, p.879) who show how one can identify product-specific search costs from conditional search moments.

to model search with learning (Dzyabura and Hauser, 2019; Hodgson and Lewis, 2020), it would be interesting to study what such learning implies for the way consumers react to new product launches. For example, if a consumer is learning their preferences during search, they may not realize that the new product matches their preferences well; they may instead focus their initial search on products that later turn out to be irrelevant. Whether such learning behavior reduces consumer surplus from new products is an intriguing question for future work.

References

Ackerberg, D. A. and M. Rysman (2005): "Unobserved product differentiation in discrete-choice models: estimating price elasticities and welfare effects," *RAND Journal of Economics*, 36, 771–789.

Bajari, P., C. L. Benkard, et al. (2001): *Discrete choice models as structural models of demand: Some economic implications of common approaches*, Graduate School of Business, Stanford University.

Berry, S., J. Levinsohn, and A. Pakes (2004): "Differentiated products demand systems from a combination of micro and macro data: The new car market," *Journal of political Economy*, 112, 68–105.

Berry, S. and A. Pakes (2007): "The pure characteristics demand model," *International Economic Review*, 48, 1193–1225.

Bresnahan, T. F. (1986): "Measuring the spillovers from technical advance: mainframe computers in financial services," *The American Economic Review*, 742–755.

Broda, C. and D. E. Weinstein (2006): "Globalization and the Gains from Variety," *The Quarterly journal of economics*, 121, 541–585.

Bronnenberg, B. J., J.-P. Dubé, M. Gentzkow, and J. M. Shapiro (2015): "Do pharmacists buy Bayer? Informed shoppers and the brand premium," *The Quarterly Journal of Economics*, 130, 1669–1726.

Bronnenberg, B. J., J. B. Kim, and C. F. Mela (2016): "Zooming in on choice: How do consumers search for cameras online?" *Marketing science*, 35, 693–712.

Brynjolfsson, E., Y. Hu, and M. D. Smith (2003): "Consumer surplus in the digital economy: Estimating the value of increased product variety at online booksellers," *Management Science*, 49, 1580–1596.

Ching, A., T. Erdem, and M. Keane (2009): "The price consideration model of brand choice," *Journal of Applied Econometrics*, 24, 393–420.

Ching, A. T., T. Erdem, and M. P. Keane (2014): "A simple method to estimate the roles of learning, inventories and category consideration in consumer choice," *Journal of choice modelling*, 13, 60–72.

Choi, M., A. Y. Dai, and K. Kim (2018): "Consumer search and price competition," *Econometrica*, 86, 1257–1281.

Christensen, C. M. (1993): "The rigid disk drive industry: A history of commercial and technological turbulence," *Business history review*, 67, 531–588.

Chung, J., P. K. Chintagunta, and S. Misra (2018): "Estimation of Sequential Search Models," *Working Paper*.

De los Santos, B. (2018): "Consumer search on the internet," *International Journal of Industrial Organization*, 58, 66–105.

De Los Santos, B. I., A. Hortacsu, and M. Wildenbeest (2012): "Testing Models of Consumer Search using Data on Web Browsing and Purchasing Behavior," *American Economic Review*, 102, 2955–2980.

Donnelly, R., A. Kanodia, and I. Morozov (2021): "The Long Tail Effect of Personalized Rankings," *Available at SSRN 3649342*.

Dzyabura, D. and J. R. Hauser (2019): "Recommending products when consumers learn their preference weights," *Marketing Science*, 38, 417–441.

Goeree, M. S. (2008): "Limited Information and Advertising in the US Personal Computer Industry," *Econometrica*, 76, 1017–1074.

Hausman, J., G. Leonard, and J. D. Zona (1994): "Competitive analysis with differenciated products," *Annales d'Economie et de Statistique*, 159–180.

Hausman, J. A. (1996): "Valuation of new goods under perfect and imperfect competition," in *The economics of new goods*, University of Chicago Press, 207–248.

Hicks, J. R. (1940): "The valuation of the social income," *Economica*, 7, 105–124.

Hodgson, C. and G. Lewis (2020): "You can lead a horse to water: Spatial learning and path dependence in consumer search," .

Honka, E. (2014): "Quantifying search and switching costs in the US auto insurance industry," *The RAND Journal of Economics*, 45, 847–884.

Honka, E. and P. K. Chintagunta (2015): "Simultaneous or sequential? search strategies in the us auto insurance industry," *Marketing Science*, 36, 21–42.

Honka, E., A. Hortaçsu, and M. A. Vitorino (2017): "Advertising, consumer awareness, and choice: Evidence from the us banking industry," *The RAND Journal of Economics*, 48, 611–646.

Honka, E., A. Hortaçsu, and M. Wildenbeest (2019): "Empirical search and consideration sets," in *Handbook of the Economics of Marketing*, Elsevier, vol. 1, 193–257.

Igami, M. (2017): "Estimating the innovators dilemma: Structural analysis of creative destruction in the hard disk drive industry, 1981–1998," *Journal of Political Economy*, 125, 798–847.

Kim, J. B., P. Albuquerque, and B. J. Bronnenberg (2010): "Online Demand under Limited Consumer Search," *Marketing Science*, 29, 1001–1023.

Link, W. A. and M. J. Eaton (2012): "On thinning of chains in MCMC," *Methods in ecology and evolution*, 3, 112–115.

Manzini, P. and M. Mariotti (2010): "Revealed preferences and boundedly rational choice procedures: an experiment," *University of St. Andrews and IZA Working Paper*, 141–167.

——— (2012): "Categorize then choose: Boundedly rational choice and welfare," *Journal of the European Economic Association*, 10, 1141–1165.

Moraga-González, J. L., Z. Sándor, and M. R. Wildenbeest (2021): "Consumer search and prices in the automobile market," .

Morozov, I., S. Seiler, X. Dong, and L. Hou (2021): "Estimation of preference heterogeneity in markets with costly search," *Marketing Science*.

Petrin, A. (2002): "Quantifying the benefits of new products: The case of the minivan," *Journal of political Economy*, 110, 705–729.

Rossi, P. E., G. M. Allenby, and R. McCulloch (2012): *Bayesian statistics and marketing*, John Wiley & Sons.

Seiler, S. (2013): "The impact of search costs on consumer behavior: A dynamic approach," *Quantitative Marketing and Economics*, 11, 155–203.

Ursu, R. M. (2018): "The Power of Rankings: Quantifying the Effect of Rankings on Online Consumer Search and Purchase Decisions," *Marketing Science*, 37, 530–552.

Weitzman, M. L. (1979): "Optimal Search for the Best Alternative," *Econometrica*, 47, 641–654.

Yang, L., O. Toubia, and M. G. De Jong (2015): "A bounded rationality model of information search and choice in preference measurement," *Journal of Marketing Research*, 52, 166–183.

Yavorsky, D., E. Honka, and K. Chen (2021): "Consumer search in the US auto industry: The role of dealership visits," *Quantitative Marketing and Economics*, 19, 1–52.

Online Appendix

A Data Construction Details

A.1 Constructing the product list

The raw Comscore data contain the URL addresses of all pages visited by each user. From these raw data, I extract the complete sequences of all Amazon product page visits of each user. This step is not trivial because several different visits to the same product page can appear with slightly different URLs depending on the reference webpage the user came from. Fortunately, the relevant product page URLs always contain the product's Amazon Standard Identification Number (ASIN), so I utilize the observed ASINs to identify all unique Amazon products that were searched by the users. In 2016, Comscore users searched in total 3,564,790 unique products on Amazon; I take this product list with corresponding ASIN codes as a starting point. I then collect data on product names, categories, attributes, brands, release dates, and daily historical prices from a third-party data service *keepa.com*.[17] This data collection step allows me to classify most of the observed searches into product categories by using product classifiers that appear on Amazon product pages. The main limitation, however, is that around 30% of products in Comscore data are no longer offered by Amazon and are not included in Keepa's dataset, so these data are missing. These are relatively unpopular products that did not manage to generate much demand on Amazon in 2016, most of which were searched only by 1-2 users. Since these products account for a small share of searches and purchases, I do not expect these missing data to significantly affect the results.

A.2 Constructing the list of hard drives

I use the category labels to construct the complete list of hard drives that were offered on Amazon in 2016. It is easy to identify hard drives in the list of all products because Amazon shows them in the category "Electronics > Computers & Accessories > Data Storage," subcategories "Internal Hard Drives" and "External Hard Drives." To avoid missing any hard drives, I additionally searched specific keywords on Amazon (e.g., "hard disk drive" or "solid state drive"), saved the list of all hard drives on the first ten pages of search results, and added these items to the assortment. The final list contains 1,774 hard drives, which I interpret as the selection of hard drives that consumers encountered when visiting Amazon in 2016. In Section 4, I simplify estimation by focusing on the 100 most frequently purchased hard drives. In doing so, I remove from the dataset all searches that correspond to hard drives outside this top 100 list of items. Since Bayesian estimation requires imputing a large $N \times J$ matrix of utilities u_{ij}, reducing the number of products helps to substantially reduce the computational burden, which in turn enables me to specify a more flexible distribution of consumer tastes α_i and β_i.

[17] Keepa offers a subscription-based API service that allows me to programmatically request data for hundreds of Amazon products each minute. The same service allows me to request data on historical prices going back to 2010-2011, which is how I collect historical prices of hard drives in the main dataset.

Since the data extracted from Keepa's database contains product attributes and prices, I use these data to construct the historical daily prices of each hard drive, time-invariant attributes x_j, and the dates on which each drive was added to Amazon's assortment. In approximately 10% of cases where attribute data are missing, I collect missing information from Amazon's product pages or informational websites about PC hardware such as *pcpartpicker.com* and *userbenchmark.com*. Similarly, I collect missing data on historical prices from specialized price-tracking websites *thetracktor.com* and *camelcamelcamel.com*. In rare cases in which none of the price-tracking websites report price data, I impute price series from purchase data for that hard drive; if no purchases are observed, I impute them from the prices of other hard drives with similar attributes (i.e., matching by brand, storage capacity, and memory type).

A.3 Computing search intensity in other categories

To construct the main search cost shifter in Section 2.4, I need to compute the number of products that consumers search in other product categories on Amazon. To this end, I first use the collected category labels to classify all Amazon products that the users searched into categories. In principle, there are several ways to do this, because Amazon provides several distinct category labels at different levels of specificity. The first two levels of Amazon's category hierarchy are too general, as using them would force me to combine many complementary products into one large category. For example, computers, computer parts, and peripherals would have to be combined into one large category of "Electronics > Computers & Accessories". The third hierarchy level provides a more granular classification but still appears rather broad. The category definitions at this level would combine laptops, desktops, and tablets into one category "Computers & Tablets", which seems overly broad. The same category definitions would also combine the equipment for ice skating, skiing, and snowboarding into a large category "Winter Sports", thus pooling products that are not close substitutes.

The fourth level of the hierarchy is more appropriate because it classifies conceptually different products into distinct categories. I therefore use this product classification for my main analysis. Visually examining the resulting categories reveals that they indeed contain highly related products that might be considered as close substitutes, although this verification step is inevitably subjective. I drop the categories with less than 100 items; these are small groups of niche products that account for less than 2% of Comscore searches. I also remove the video and music-streaming categories "Movies & TV" and "Digital Music," as it is unclear how to properly define and measure consumer search in such categories. Overall, this process enables me to classify all 3,564,790 unique Amazon products in Comscore data into 2,165 product categories. To illustrate the granularity of the resulting categories, Table 7 shows several examples of selected categories from four large product classes.

Having classified products into categories, I compute how many items each user searched in each product category in 2016. I then regress the resulting number of searches per category on user fixed effects as well as on category fixed effects in order to partial out the impact of category-specific

	Searched any SSD	Number of searches All	SSDs	HDDs	Purch. cond. on search Low Price	Low TB
Tech-Savvy Indicator	0.181***	0.754***	0.633***	0.121	-0.116	-0.131
	(0.032)	(0.205)	(0.145)	(0.158)	(0.146)	(0.149)
Searches per Session	0.023***	0.280***	0.127***	0.153***	-0.005	-0.021
(other categories)	(0.004)	(0.047)	(0.018)	(0.043)	(0.011)	(0.015)
Time on Social Media	0.043***	-0.054	0.073	-0.128**	0.022	0.050
	(0.016)	(0.056)	(0.047)	(0.050)	(0.045)	(0.053)
Time on Video & Music	-0.059***	0.197	-0.179***	0.376**	0.121**	0.068
	(0.018)	(0.140)	(0.025)	(0.147)	(0.059)	(0.090)
Time on Adult Sites	0.010**	0.040*	0.026	0.014	-0.012	-0.002
	(0.004)	(0.021)	(0.016)	(0.016)	(0.011)	(0.009)
High Total Time Online	0.027	0.384***	0.119**	0.265***	0.029	0.026
(above median)	(0.022)	(0.096)	(0.057)	(0.091)	(0.053)	(0.043)
Demographic controls	Yes	Yes	Yes	Yes	Yes	Yes
Observations	2,422	2,422	2,422	2,422	222	222
R-squared	0.048	0.097	0.078	0.040	0.019	0.032

Table 6: **Correlations of consideration and search cost shifters with searches and purchases.** Source: Comscore, Web-Behavior Panel, January 2016-December 2016, U.S.

search costs. Figure 6 shows that the estimated user fixed effects show significant variation across users in the main sample. I interpret these user fixed effects as this user's expected number of searches relative to the average number of searches in a given category. Users with low estimated fixed effects search fewer products compared to the average search behavior in a specific category, whereas users with high fixed effects search more than average. I use these estimated user fixed effects as the main search cost shifter (see Section 2.4 for the related discussion).

A.4 Constructing consideration shifters and time use variables

Another search cost shifter in Section 2.4 requires me to classify the websites visited by Comscore consumers into *PC hardware websites*, *gaming and esports websites*, *social media*, *adult entertainment*, and *video and music streaming*. To remove arbitrary decisions from this classification task, I use Google Search results as a way to classify the visited pages. To this end, I use keywords relevant to each class of websites as search queries, and I save the first five pages of search results for each query. For example, the queries I use to identify PC hardware websites are "computer hardware," "computer hardware news," "PC hardware," and "PC hardware news." Similarly, for gaming-related websites, I use queries "gaming news," "video game news." and "esports news." Having obtained the list of websites displayed by Google search, I then manually remove online stores (e.g., Amazon and NewEgg) as well as non-specialized news outlets (e.g., Wall Street Journal and Bloomberg). Table 8 shows the resulting list of websites, which I use to identify tech-savvy users. I repeat this process for other website classes and merge the resulting website list with the

No.	Product Class	Category Name	No. Products
1	Electronics	Bags, Cases & Sleeves	9525
2	Electronics	Electronics Features	5423
3	Electronics	Internal Components	4770
4	Electronics	Cables & Interconnects	3790
5	Electronics	Laptops	3721
6	Electronics	Earbud Headphones	3034
7	Electronics	'MP3 & MP4 Player Accessories	2558
8	Electronics	Keyboards, Mice & Accessories	2289
9	Electronics	Car Audio	2261
10	Electronics	Cables & Accessories	2212
11	Beauty & Personal Care	Cleansers	3046
12	Beauty & Personal Care	Moisturizers	2357
13	Beauty & Personal Care	Treatments & Masks	2089
14	Beauty & Personal Care	Creams & Moisturizers	1774
15	Beauty & Personal Care	Shampoos	1237
16	Beauty & Personal Care	Wigs	1178
17	Beauty & Personal Care	Lipstick	1158
18	Beauty & Personal Care	Nail Polish	1147
19	Beauty & Personal Care	Hair Extensions	1126
20	Beauty & Personal Care	Foundation	1051
21	Books	Subjects	21096
22	Books	Genre Fiction	13554
23	Books	Erotica	5804
24	Books	Literature & Fiction	5565
25	Books	Contemporary	5197
26	Books	Mystery	4113
27	Books	United States	3988
28	Books	Paranormal	2928
29	Books	Historical	2443
30	Books	Action & Adventure	2440
31	Toys & Games	Collectible Card Games	7260
32	Toys & Games	Toy Figures & Playsets	4300
33	Toys & Games	Action Figures	4231
34	Toys & Games	Dolls	3039
35	Toys & Games	Board Games	2591
36	Toys & Games	Building Sets	2529
37	Toys & Games	Stuffed Animals & Teddy Bears	1768
38	Toys & Games	Playsets	1470
39	Toys & Games	Jigsaw Puzzles	1453
40	Toys & Games	Costumes	1450

Table 7: **Examples of product categories used for computing individual-level search intensity.** These categories correspond to the fourth level of product hierarchy defined by category labels on Amazon's product pages. The second column specifies a "product class" which corresponds to the first level of hierarchy. The third column reports category names for the largest categories within several different product classes. The last column shows the number of products classified into each category. Source: Comscore, Web-Behavior Panel, January 2016-December 2016, U.S.

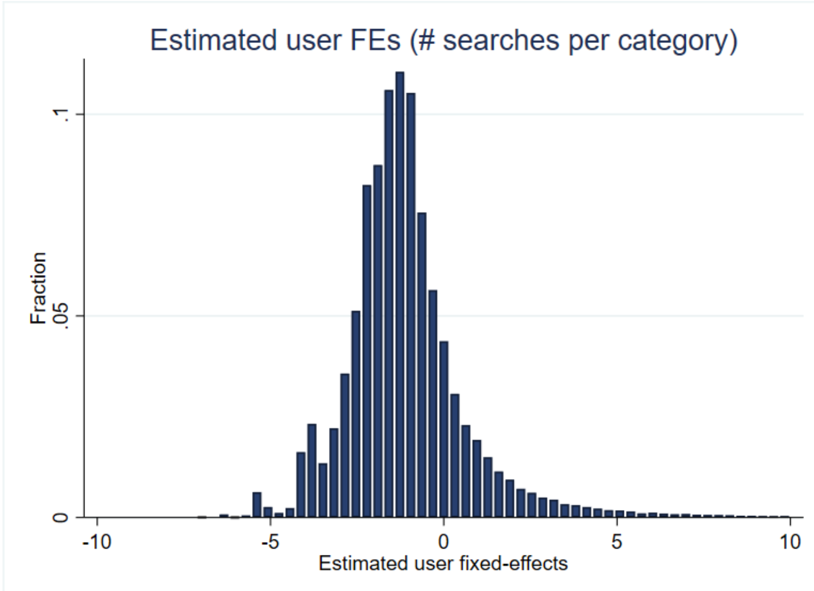

Figure 6: **Estimated user fixed effects.** The graph plots the distribution of estimated user fixed effects, obtained by regressing the number of searches per category on user fixed effects and category fixed effects. I utilize these estimated user fixed effects as a main search cost shifter (see Section 2.4).

main Comscore dataset. This process enables me to classify all web pages in the raw Comscore logs and to compute how much time users spend on various online activities, e.g., reading about gaming or computer hardware, browsing social media, and so on.

Users may not be willing to spend much time searching if several other tasks compete for their limited time and attention. Table 6 shows that users who spend a lot of time online (i.e., above the median in the sample) search a larger number of hard drives. However, the total time spent online does not affect the probability of searching at least one SSD or the purchase probabilities conditional on search, suggesting that it is not correlated with consideration or preferences. Other time-use variables seem to correlate with the key outcome variables in arbitrary ways, and more importantly, they correlate with the attributes of purchased hard drives (e.g., see "Video & Music" websites in column 5). For this reason, I do not use these variables as shifters and use only "searches per category" and "total time online" as shifters of search costs.

A.5 Selecting relevant product page visits

When defining search sessions in Section 2, I need to identify all searches related to a specific choice occasion. In the data, I see that over 80% of all product page visits occur within a week prior to a purchase. Based on this observation, for users with purchases, I select all product page visits that occurred within a week before the day of the purchase. Selecting relevant visits is more challenging for users without purchases because it is difficult to determine when they finished their search. Operationally, for these users, I choose the week in 2016 on which they visited the largest number of product pages in the hard drive category. I then interpret this week as a search session, and

Websites classified as *'computer hardware websites'*:
Explainingcomputers.com, juicedsystems.com, malabs.com, microcenter.com, ecomputerz.com, extremetech.com, gamersnexus.net, hardwareasylum.com, hardwarecanucks.com, insight.com, kitguru.net, linustechtips.com, modders-inc.com, pcgameshardware.de, play3r.net, pureoverclock.com, techpowerup.com, techradar.com, techspot.com, thinkcomputers.org, turnhardware.net, wccftech.com, anandtech.com, cgdirector.com, computeruniverse.net, computerweekly.com, computerworld.com, crucial.com, crucial.in, digitalstorm.com, lifewire.com, maketecheasier.com, nzxt.com, openhardwaremonitor.org, pchardwarehelp.com, pchardwarelinks.com, pcpartpicker.com, pcper.com, techguided.com, windowscentral.com, bit-tech.net, computerbild.de, eweek.com, extremeoverclocking.com, fractal-design.com, game-debate.com, geeksultd.com, guru3d.com, hardwarebbq.com, hardwaretimes.com, hexus.net, hothardware.com, pc-builds.com, pchardware.co.uk, pchardwarerefresh.com.au, pcmag.com, pcmasters.de, pc-specs.com, pcworld.com, performancepsu.com, tech4gamers.com, techadvisor.com, techarx.com, technewstoday.com, techspy.com, thepcenthusiast.com, tomshardware.com, tweaktown.com, vortez.net, wepc.com.
Websites classified as *'gaming and esports websites'*:
Alliedesports.gg, dailyesports.gg, dotesports.com, escharts.com, escorenews.com, esports.com, esports.net, esports.swac.org, esportsentertainmentgroup.com, esportsinsider.com, esportsjunkie.com, esportsobserver.com, esportsph.com, esportsreporter.com, esportstales.com, esportstalk.com, esportstechnologies.com, esportznetwork.com, gamerworldnews.com, gears.gg, gfinityesports.com, ggrecon.com, givemesport.com, global-esports.news, gosugamers.net, halowaypoint.com, lolesports.com, magic.gg, news.unikrn.com, siege.gg, snowballesports.com, sooneresports.org, talkesport.com, thegamer.com, thetimesofesports.com, tl.net, vlr.gg, weplay.tv, win.gg, xbox.com, amazongames.com, g2a.com, game.intel.com, gameinformer.com, gamerant.com, gamesradar.com, gaminglyfe.com, happygamer.com, kiriludos.com, kotaku.com, news.xbox.com, pcgamer.com, pcgamesn.com, riotgames.com, rockpapershotgun.com, tabletopgamingnews.com, thegamerspost.com, twinfinite.net, twitch.tv, videogameschronicle.com, destructoid.com, engadget.com, gamepur.com, gameranx.com, gamesindustry.biz, gamespot.com, gamestop.com, gamingbolt.com, mcvuk.com, mmorpg.com, play.google.com, polygon.com, shacknews.com, theverge.com.

Table 8: **List of PC hardware and gaming websites for identifying tech-savvy users.**

I interpret all product pages visited in that week as relevant searches. This definition of search sessions seems reasonable because product page visits are relatively concentrated. As a result, defining a search session as being one week long only removes about 15%-20% of all searches from the dataset.

A.6 Additional data summary: search persistence

Users focus their search on hard drives that share similar attributes, as I show in Table 9. Each entry in this table is the conditional probability that a user transitions from searching the row brand to searching the column brand, where each transition is between two subsequent product page visits. For example, after having searched one Crucial hard drive, the user then searches another Crucial hard drive with a probability 31.8%, a Samsung hard drive with a probability 22.7%, or a Seagate hard drive with a probability 13.6%. Observe that the diagonal entries are several times higher than unconditional search probabilities in the last row of the table. Hence,

Brand	Brand (next searched hard drive)							
(current drive)	Crucial	Samsung	SanDisk	Seagate	Toshiba	Transcend	WD	Other
Crucial	**31.8**	22.7	0.0	13.6	0.0	0.0	18.2	13.6
Samsung	9.8	**41.3**	6.5	14.1	0.0	0.0	14.1	14.1
SanDisk	3.9	7.7	**53.9**	11.5	3.9	0.0	3.9	15.4
Seagate	1.3	2.9	1.1	**62.2**	3.2	1.9	21.2	6.4
Toshiba	0.0	3.0	1.5	16.4	**46.3**	1.5	22.4	9.0
Transcend	4.4	0.0	0.0	26.1	4.4	**26.1**	30.4	8.7
Western Digital	0.5	6.3	0.0	21.2	3.3	1.6	**60.6**	6.5
Other	1.7	6.0	4.3	24.8	1.7	0.9	22.2	**38.5**
Uncond. prob.	2.5	8.1	2.7	34.6	5.4	1.9	33.8	11.1

Table 9: **Transition matrix: persistence of consumer search with respect to different brands.** Source: Comscore Web-Behavior Panel, January 2016-December 2016, U.S.

consumers systematically focus their search on the hard drives of specific brands, which generates substantially stronger search persistence than would be implied by random search.

Similar observations can be made about other hard drive characteristics. Of particular interest are transition probabilities for HDDs and SSDs. Around 90.4% of HDD searches are followed by another HDD search, and around 62.4% of SSD searches are followed by another SSD search. As before, high transition probabilities suggest that consumers focus on searching hard drives from the same category. However, in 37.3% of cases, consumers switch from searching an SSD to searching an HDD, which indicates that hard drives of these two types are perceived as substitutes.

Figure 7 further illustrates search persistence by presenting it in two dimensions: brand and type (HDD vs SSD). Each cell in the matrix shows the probability that a user whose first search was a hard drive from the "row" category also searched at least one hard drive from the "column" category within the same session. Darker color implies a higher co-search probability. The figure appears consistent with the search persistence documented in Table 9. Consumers focus on searching hard drives of a specific brand and type (e.g., Seagate SSDs), and they rarely search hard drives of other brands or types. Nevertheless,there are some high off-diagonal probabilities in the northeast and the southwest regions of Figure 7, suggesting that some consumers have a strong preference for hard drive types (e.g., strongly prefer HDDs) but search across brands. Additionally, some co-search probabilities are large in the southeast region of the figure, implying some degree of substitution between SSDs and HDDs. These observations motivate me to specify a choice model that can flexibly capture this substitution, modeling it through a combination of consideration decisions and heterogeneous preferences.

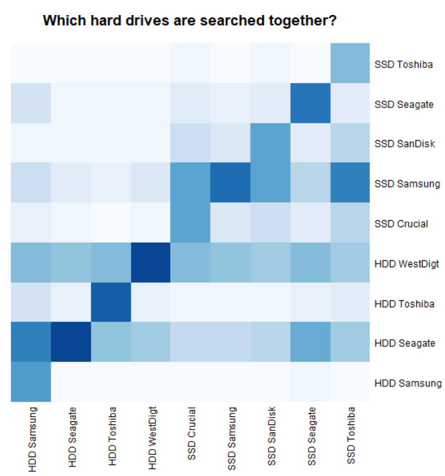

Figure 7: **The attributes of hard drives that are often searched together.** This matrix illustrates which brands and types of hard drives are often searched together within the same search session. Each cell describes the probability that a user whose first search was hard drive from the "row" category also later searched at least one hard drive from the "column" category in the same search session. Darker color implies a higher conditional probability. The computed probabilities are based on all 1,774 hard drives in the Comscore dataset. Source: Comscore, Web-Behavior Panel, January 2016-December 2016, U.S.

B Computing Reservation Utilities

B.1 Decomposition of reservation utilities z_{ij}

Lemma 1 Suppose taste shocks ε_{ij} follow a continuous atomless distribution with CDF $F_\varepsilon(\cdot)$. The following statements then hold for the reservation utility z_{ij} defined in Section 3.2:

1. The reservation utility z_{ij} can be expressed as $z_{ij} = \delta_{ij} + \xi(c_i)$, where $\xi(c_i)$ is a decreasing function in the search cost c_i defined as the unique solution to the equation $\int_{\varepsilon_z}(1-F_\varepsilon(\varepsilon))d\varepsilon = c_i$.

2. If $F_\varepsilon(\cdot)$ belongs to the location-scale family of distributions, the value of $\xi(c_i)$ can be computed as follows. Let $\Gamma(\psi; c_i)$ be a contraction mapping $\Gamma(\psi; c_i) = \int_\xi(\varepsilon - \psi)dF_\varepsilon(\varepsilon) + \psi - (c_i/\sigma_\varepsilon)$. Then, $\xi(c_i)$ can be expressed as $\xi(c_i) = \sigma_\varepsilon \cdot \psi(c_i)$, where $\psi(c_i)$ is defined as the unique solution to the equation $\Gamma(\psi; c_i) = \psi$.

3. When taste shocks ε_{ij} are normally distributed, the contraction mapping can be expressed as $\Gamma(\psi; c_i) = \psi \cdot \Phi(\psi) + \phi(\psi) - (c_i/\sigma_\varepsilon)$, where $\phi(\psi)$ and $\Phi(\psi)$ are PDF and CDF of the standard normal distribution.

Proof (1) This result appears in Choi et al. (2018) without proof. To see why the result is true, integrate the expression for the reservation utility by parts to obtain $\int_z(1 - F_u(u_{ij}))du_{ij} = c_i$, where $F_u(\cdot)$ is the CDF of u_{ij}. The result follows immediately after changing the variable of integration to $\varepsilon_{ij} = u_{ij} - \delta_{ij}$ and defining $\xi = z_{ij} - \delta_{ij}$. Also note that $\xi(c_i)$ is decreasing, because $\partial\xi(c_i)/\partial c_i = -1/(1 - F(\xi)) < 0$ by the implicit function theorem.

(2) Assume $F_\varepsilon(\cdot)$ belongs to the location-scale family so that $F_u(u_{ij}) = F_\varepsilon((u_{ij} - \delta_{ij})/\sigma_\varepsilon)$ and $f_u(u_{ij}) = (1/\sigma_\varepsilon)f_\varepsilon((u_{ij} - \delta_{ij})/\sigma_\varepsilon)$. Changing the variable of integration in the expression for the reservation utility to $\varepsilon_{ij} = (u_{ij} - \delta_{ij})/\sigma_\varepsilon$ and letting $\psi_{ij} = (z_{ij} - \delta_{ij})/\sigma_\varepsilon$, we obtain $\int_{\xi_{ij}}(\varepsilon_{ij} - \psi_{ij})dF(\varepsilon_{ij}) - (c_i/\sigma_\varepsilon) = 0$. The mapping $\Gamma(\psi; c_i)$ in the lemma can be constructed by adding ψ_{ij} to both sides of this equation. In addition, note $\Gamma(\psi; c_i)$ is a contraction mapping, because $\Gamma'(\psi; c_i) = -(1 - F(\psi)) + 1 = F(\psi) \in (0, 1)$ for any finite value of ψ.

(3) Assuming $\varepsilon_{ij} \sim N(0, \sigma_\varepsilon^2)$, I can rewrite the first part of the contraction mapping as $\int_\xi(\varepsilon - \psi)dF_\varepsilon(\varepsilon) = Pr(\varepsilon \geq \psi)E(\varepsilon - \psi|\varepsilon \geq \psi)$, which, under the normality assumption, simplifies to $\int_\xi(\varepsilon - \psi)dF_\varepsilon(\varepsilon) = \phi(\psi) - (1 - \Phi(\varepsilon)) \cdot \psi$. Plugging this result back into the expression for $\Gamma(\psi; c_i)$ yields the expression in the lemma. I use this expression to compute the values of $\psi(c_i)$ and $\xi(c_i)$. Also note that the function $\psi(c_i)$ can be inverted analytically as $\psi^{-1}(x) = \sigma_\varepsilon \cdot (\phi(x) - x \cdot (1 - \Phi(x)))$, which I extensively use in the counterfactual analyses in Section 4. Kim et al. (2010) use a similar result to compute reservation utilities within their estimation algorithm.

C Details of the MCMC Sampler

Setup. In this section, I further explain the proposed MCMC sampler by presenting all relevant posterior distributions and stating the priors used for estimation. As mentioned in Section 3.3, the goal of estimation is to recover the hyperparameters that determine the distribution of user types λ_i. The type λ_i of each user includes the price coefficient α_i, tastes $\beta_{1i}, \dots, \beta_{Ki}$, search propensities ξ_i, and consideration parameters γ_i^{SSD} and γ_i^{HHD}. Each element of the type vector, λ_i^k, is independent from other elements and follows the normal distribution such that:

$$\lambda_i^k \sim N(\pi_k' w_i^k, \sigma_k^2) \tag{5}$$

where w_i^k is a column-vector of L_k user characteristics included in the specification of the parameter λ_i^k (always includes a constant), π_k is a column-vector of L_k regression coefficients, and σ_k^2 is the variance capturing unobserved heterogeneity. This formulation allows me to handle each element of λ_i independently by estimating a separate Bayesian regression of that parameter on relevant user characteristics z_i. In what follows, I derive an MCMC sampler that takes draws of variables λ_i while respecting all search and choice inequalities. Conditional on these draws, the sampler then projects the resulting draws of types onto user characteristics w_i to draw the regression coefficients π_k and variances σ_k^2. Although I present the MCMC sampler for the sequential search model, the method can be applied to other search models, e.g., the simultaneous search model from Honka (2014) or a consideration set model from Goeree (2008), with only minor modifications.

Choice inequalities. I simplify the system of restrictions in Section 3.3 as follows. Given the decomposition of reservation utilities $z_{ij} = \delta_{ij} + \xi_i$ from Lemma 1, the optimal search strategy is to search in order of decreasing pre-search utilities δ_{ij}. I drop several inequalities from the system as redundant. For example, it suffices to ensure that $\max(u_{i,1}, \dots, u_{i,K_i-1}) \leq z_{i,K_i}$ to verify that all other continuation inequalities hold as well. These simplifications leave us with the following set of inequalities:

$$\begin{array}{lll}
j \in C_i \iff q_{ij} \geq 0 & \forall j & \text{(consideration)} \\
\delta_{i,1} \geq \delta_{i,2} \geq \cdots \geq \delta_{i,K_i} \geq \delta_{i,j} & \forall j \in C_i \setminus S_i & \text{(search order)} \\
\max(u_{i,0}, u_{i,1}, \dots, u_{i,K_i-1}) \leq \delta_{i,K_i} + \xi_i & & \text{(continuation)} \\
\max(u_{i,0}, u_{i,1}, \dots, u_{i,K_i}) \geq \delta_{i,j} + \xi_i & \forall j \in C_i \setminus S_i & \text{(stopping)} \\
u_{i,y_i} \geq u_{i,j} & \forall j \in S_i \cup \{0\} & \text{(purchase)}
\end{array}$$

Importantly, all these inequalities are linear in pre-search utilities δ_{ij} and search propensities ξ_i.

This property proves critical for estimation as it helps me derive simple upper and lower bounds for unobserved parameters of the model. Without this property, I would need to resort to an inefficient acceptance-rejection sampling algorithm that rejects all draws that do not satisfy the inequality conditions.

Priors. Bayesian methods require me to specify the prior distributions of parameters π_k and σ_k^2 for each element of the vector of types λ_i. Following Rossi et al. (2012, pp. 21-27), I assume conjugate priors such that the variance σ_k^2 follows the inverse gamma distribution so that $\sigma_k^2 \sim IG(v_0/2, v_0 s_0^2/2)$, and the vector π_k follows the normal distribution conditional on the variance σ_k^2 so that $\pi_k|\sigma_k^2 \sim N(\bar{\pi}_k, \sigma_k^2 A^{-1})$. I assume weakly informative priors for coefficients π_k by setting $\bar{\pi}_k = 0$ and $A^{-1} = 10 \cdot I_{L_k}$. In practice, I found that changing these priors makes little difference for the resulting posterior distributions because search and purchase data contain rich information about coefficients π_k. Additionally, I also assume diffuse priors for variances σ_k^2 by setting $v_0 = L_k$ and $s_0^2 = 0.01$. Doing so achieves two goals: it puts a sufficient mass on small values of variance σ_k^2, and it shrinks unobserved heterogeneity estimates toward zero just enough to prevent overfitting. By contrast, choosing a larger prior variance, such as $s_0^2 = 1.0$, leads to implausibly large variance estimates, which reflects the fact that unobserved heterogeneity is not well identified without "true" panel data (see Section 3.4. for discussion). Lastly, for the variance of the pre-search shocks η_{ij} , I assume the inverse gamma prior such that $\sigma_\eta^2 \sim IG(v_0/2, v_0 s_0^2/2)$ with $v_0 = 0.01 \cdot NK$ and $s_0^2 = 1.0$. This prior ensures that the shocks η_{ij} do not a priori dominate the utility u_{ij}.

Additional notation. It helps to introduce several notation shortcuts. I use δ_i and u_i to denote vectors that stack product-specific utilities for user i. Instead of referring to tastes as α_i and β_i, I stack both parameters into one taste vector $\theta_i = (\alpha_i, \beta_i')'$ and all hard drive attributes in the same vector $\tilde{x}_{ij} = (p_{j,t(i)}, x_j')'$. I also occasionally refer to $\bar{\delta}_i = \max_{m \in C_i \setminus S_i} \delta_{i,m}$ as the highest pre-search utility outside the search set S_i, $\underline{\delta_i} = \min_{m \in S_i} \delta_{i,m}$ as the lowest pre-search utility inside the search set S_i, and $\bar{z}_i = \max_{j \in C_i \setminus S_i} \{z_{ij}\}$ as the highest reservation utility in S_i. Similarly $u_i^* = \max_{j \in S_i} u_{ij}$ captures the highest utility among the searched hard drives and $u_{i,(j)}^* = \max_{j \in S_i \setminus \{j\}} u_{ij}$ is the highest utility among the searched hard drives excluding product j itself (both maxima include the outside option utility u_{i0}). Finally, I let $D_i = \{S_i, y_i\}$ denote the observed searches and purchases of user i.

Step I. Drawing pre-search utilities δ_i. The first step draws utilities δ_i given the type λ_i, search propensities ξ_i, utilities u_i, and data D_i. I draw δ_{ij} sequentially for each hard drive, each time conditioning on the vector of utilities for all other hard drives except j, denoted as the leave-one-out vector $\delta_i^{(j)}$. Given the model structure, utilities δ_{ij} are conditionally normal so that $\delta_{ij}|\theta_i \sim N(\tilde{x}_j'\theta_i, \sigma_\eta^2)$ and $u_{ij}|\delta_{ij} \sim N(\delta_{ij}, \sigma_\varepsilon^2)$. It follows that δ_{ij} conditional on u_{ij}, $\tilde{x}_j$, and λ_i is also normal when I do not condition on the observed choices D_i. When I do condition on D_i, the distribution of δ_{ij} becomes truncated normal:

$$\delta_{ij}|\delta_i^{(j)}, u_i, \lambda_i, \xi_i, \tilde{x}_j, D_i \sim trN\left(\delta_{ij}^*, (\sigma^*)^2, \underline{\delta}_{ij}, \bar{\delta}_{ij}\right),$$

where $\delta^*_{ij} = (u_{ij}\sigma^2_\eta + (\tilde{x}'_j\lambda_i)\sigma^2_\varepsilon)/(\sigma^2_\eta + \sigma^2_\varepsilon)$ and $\sigma^* = \sigma_\eta\sigma_\varepsilon/\sqrt{\sigma^2_\eta + \sigma^2_\varepsilon}$. Here, the truncation points $\underline{\delta}_{ij}$ and $\bar{\delta}_{ij}$ are determined by choice inequalities summarized at the beginning of this section. For example, when user i searches strictly more than one but fewer than J_i hard drives ($1 < K_i < J_i$), I truncate utilities δ_{ij} as follows:

$$\begin{aligned}
&\delta_{i,j} \geq \delta_{i,j+1} && for \quad j = 1\\
&\delta_{i,j-1} \geq \delta_{i,j} \geq \delta_{i,j+1} && for \quad j = 2, \dots, K_i - 1\\
&\delta_{i,j-1} \geq \delta_{i,j} \geq \max\left\{\bar{\delta}_i, u^*_{i,(j)} - \xi_i\right\} && for \quad j = K_i\\
&\min\left\{\underline{\delta}_i, u^*_i - \xi_i\right\} \geq \delta_{i,j} && for \quad j \in C_i \setminus S_i,
\end{aligned}$$

The truncation points can be constructed in a similar way for other cases. When the user searches only one hard drive ($K_i = 1$), I drop the continuation inequality but keep all other inequalities; and when the user searches all hard drives ($K_i = J_i$), I drop the stopping inequality but keep all other inequalities. Finally, for hard drives outside the consideration set C_i, I take draws from the unconstrained distribution $\delta_{ij} \sim N(\tilde{x}'_j\theta_i, \sigma^2_\eta)$, because the observed choices convey no information about these pre-search utilities. While drawing utilities δ_{ij} outside the consideration set C_i may seem unhelpful, these draws serve as inputs to imputing the consideration sets.

Step II. Drawing post-search utilities u_i. The second step draws utilities u_i given the pre-search utilities δ_i, search propensities ξ_i, and data D_i. As in the previous step, I take these draws sequentially for each hard drive j while conditioning on the values of utilities of all other hard drives. Recall that u_{ij} is conditionally normal such that $u_{ij}|\delta_{ij} \sim N(\delta_{ij}, \sigma^2_\varepsilon)$. I therefore draw u_{ij} from the truncated normal distribution:

$$u_{ij}|u_{i,(j)}, \delta_i, \xi_i, D_i \sim trN(\delta_{ij}, \sigma^2_\varepsilon, \underline{u}_{ij}, \bar{u}_{ij}),$$

The bounds $\underline{u}_{ij}$ and $\bar{u}_{ij}$ are determined by the choice inequalities derived earlier. When user i searches strictly more than one but fewer than J_i hard drives ($1 < K_i < J_i$), the utilities u_{ij} are truncated so that

$$\begin{aligned}
\max\left\{u^*_{i,(j)}, \bar{z}_i\right\} \leq u_{ij} \leq z_{i,K_i} \quad if \quad j < K_i, \quad y_i = j\\
u_{ij} \leq \min\left\{u^*_{i,(j)}, z_{i,K_i}\right\} \quad if \quad j < K_i, \quad y_i \neq j,
\end{aligned}$$

The truncation bounds for the last searched hard drive, $j = K_i$, are somewhat different and can be computed as

$$\begin{aligned}
\max\left\{u^*_{i,(j)}, \bar{z}_i\right\} \leq u_{ij} \quad if \quad j = K_i, \quad y_i = j\\
u_{ij} \leq u^*_{i,(j)} \quad if \quad j = K_i, \quad y_i \neq j.
\end{aligned}$$

I derive the truncation points for the remaining cases ($K_i = 1$ and $K_i = J_i$) by removing continuation and stopping conditions from the system of search inequalities. The utilities of unsearched but considered hard drives are left unrestricted, so I draw them from the normal distribution $u_{ij} \sim N(\delta_{ij}, \sigma_\varepsilon^2)$ for $j \in C_i \setminus S_i$. Finally, since hard drives outside the consideration set C_i never get searched, the observed choices impose no restrictions on their utilities u_{ij}. I therefore do not include these draws in the sampling algorithm and only draw u_{ij} for hard drives in the consideration set C_i.

Step III. Drawing individual preferences θ_i. The third step draws preferences θ_i given utilities δ_i, attributes $\tilde{x}_j$, and heterogeneity parameters π_k and σ_k^2. Recall that each element of θ_i is distributed as $\theta_i^k \sim N(\pi_k' w_i^k, \sigma_k^2)$. Let Π be the matrix of regression coefficients formed by vectors π_k, and let $\Omega = \text{diag}(\sigma_1^2, \ldots, \sigma_K^2)$ be the covariance matrix such that the tastes follow the conditional distribution $\theta_i | \Pi, \Omega, w_i^p \sim N(\Pi w_i^p, \Omega)$, where w_i^p is the complete vector of taste shifters. Pre-search utilities are conditionally normal such that $\delta_i | \theta_i \sim N(X\theta_i, \sigma_\eta^2 \cdot I_J)$ where X is the matrix of attributes in which the j-th row is $\tilde{x}_j'$. Let $\tilde{\delta}_i = \delta_i / \sigma_\eta$ be the vector of normalized pre-search utilities, and let $\tilde{X} = X/\sigma_\eta$ be the matrix of normalized hard drive attributes. Then, the vector θ_i is conditionally normal and can be drawn from the distribution:

$$\theta_i | \delta_i, X, \Pi, \Omega \sim N\left(\tilde{\lambda_i}, (\tilde{X}'\tilde{X} + \Omega^{-1})^{-1}\right),$$

where $\tilde{\lambda}_i = (\tilde{X}'\tilde{X} + \Omega^{-1})^{-1}(\tilde{X}'\tilde{\delta}_i + \Omega^{-1}(\Pi w_i^p))$ is a vector of regression coefficients from the Bayesian regression of utilities $\tilde{\delta}_i$ on hard drive attributes $\tilde{X}$.

Step IV. Drawing individual search propensities ξ_i. The fourth step draws search propensities ξ_i from the following truncated normal distribution:

$$\xi_i | u_i, \delta_i, \pi_s, \sigma_s^2 \sim trN\left(\pi_s' w_i^s, \sigma_s^2, \underline{\xi(i)}, \overline{\xi(i)}\right)$$

where $\underline{\xi(i)}$ and $\overline{\xi(i)}$ are truncation bounds determined by the choice inequalities:

$$u_{i,(K_i)}^* - \delta_{i,K_i} = \underline{\xi(i)} \leq \xi_i \leq \overline{\xi(i)} = u_i^* - \bar{\delta}_i$$

where u_i^* is the highest utility in the search set S_i, and $u_{i,(K_i)}^*$ is the same maximum utility but excluding the last searched product K_i. These bounds are intuitive: the search propensity ξ_i cannot be too low, or otherwise the user would not search hard drives $2, \ldots, K_i$ after searching hard drive 1. At the same time, the search propensity ξ_i cannot be too high, otherwise the user would search other alternatives beyond the hard drive K_i. The derived bounds apply to all cases where K_i is strictly between 1 and $|C_i|$ (search more than one item but not the whole consideration set). When $K_i = 1$, one needs to remove the lower bound $\underline{\xi(i)}$, because the search propensity can in principle be infinitely low. When $K_i = |C_i|$, one needs to remove the upper bound $\overline{\xi(i)}$, because the search propensity can be infinitely high to have induced the user to search the entire consideration set C_i.

Step V. Drawing individual consideration parameters γ_i. The fifth step imputes individual consideration sets C_i and draws consideration parameters γ_i^{HDD} and γ_i^{SSD}. I start by imputing the consideration sets C_i. Let $Q_{ij} = 1(j \in C_i)$ be the indicator that hard drive j is included into user i's consideration set. For hard drives that user i actually searched ($j \in S_i$), I impute $Q_{ij} = 1$ as in this model users cannot search products that they do not consider. For an unsearched hard drive ($j \notin S_i$), I impute $Q_{ij} = 0$ if its reservation utility z_{ij} exceeds the maximum utility u_i^*, or if its pre-search utility δ_{ij} exceeds the lowest pre-search utility $\underline{\delta_i}$ of hard drives in the search set. This is because, had this hard drive been considered, the user would have included it in the search set S_i according to the optimal search rule. Finally, for unsearched hard drives that do not violate optimal search inequalities, I draw Q_{ij} from the distribution $Q_{ij} \sim \text{Bernoulli}(pr_{ij})$ where pr_{ij} is the probability that the consideration index $q_{ij} = d_j^{HDD}\gamma_i^{HDD} + d_j^{SSD}\gamma_i^{SSD} + \mu_{ij}$ exceeds zero.

The consideration parameters γ_i are then drawn in two steps. I first draw consideration indices q_{ij} from the distribution $q_{ij} \sim N(d_j^{HDD}\gamma_i^{HDD} + d_j^{SSD}\gamma_i^{SSD}, \sigma_\mu^2)$ truncated with zero from below for $j \in C_i$ and with zero from above for $j \notin C_i$. Once indices q_{ij} are drawn, I draw consideration parameters γ_i^{HDD} and γ_i^{SSD} by estimating a Bayesian regression of q_{ij} on the HDD and SSD indicators.

Step VI. Drawing hyperparameters π_k and σ_k^2. The sixth step imputes hyperparameters π_k and σ_k^2 for each element in the type vector λ_i. Given the assumed conjugate priors, the posterior distribution of σ_k^2 is the inverse gamma, and the posterior distribution of π_k conditional on σ_k^2 is normal, such that:

$$\pi_k|\sigma_k^2, \Lambda_k, w_k \sim N\left((W_k'W_k + A)^{-1}(W_k\Lambda_k + A\bar{\pi}_k), \sigma_k^2 \cdot (W_k'W_k + A)^{-1}\right)$$

$$\sigma_k^2|\Lambda_k, w_k \sim IG\left(\frac{v_1}{2}, \frac{v_1 s_1^2}{2}\right)$$

where Λ_k is a vector that stacks the k-th element of type λ_i across users (e.g., a vector of all user's price sensitivities α_i of length N), and W_k is a matrix of user characteristics in which the i-th row is w_i^k. Here the updated posterior parameters are $v_1 = L_k + N$, $s_1^2 = (L_k s_0^2 + ns^2)/(L_k + n)$, where $s^2 = (\Lambda_k - W_k\hat{\pi}_k)'(\Lambda_k - W_k\hat{\pi}_k)/(n - L_k)$ and $\hat{\pi}_k = (W_k'W_k)^{-1}W_k\Lambda_k$.

Step VII. Drawing the variance σ_η^2. The last step draws the variance σ_η^2 of pre-search shocks η_{ij}. Define residuals r_{ij} as $r_{ij} = \delta_{ij} - x_j'\beta_i + \alpha_i p_{j,t(i)}$. Then the posterior distribution of the variance σ_η^2 conditional on preferences θ_i is the inverse gamma distribution such that $\sigma_\eta^2|\theta_i, X \sim IG(v_1/2, v_1 s_1^2/2)$ where $s_1^2 = (v_0 s_0^2 + NJs^2)/(v_0 + NJ)$, $s^2 = (1/NJ)\sum_{ij} r_{ij}$, $v_1 = v0 + NJ$, and $v_0 = 0.01 \cdot NK$.

Implementation Details. The MCMC sampler successively draws from relevant conditional distributions, producing a sequence of posterior draws. In practice, Monte Carlo simulations show that the chain of draws converges to a stable posterior distribution after about 2,000-3,000 draws from arbitrary initial values of parameters. One could in principle speed up convergence by choosing better initial values, e.g., by calibrating some key parameters (e.g., search propensity and the value of the outside option) via a simple method of moments estimation. In practice, however, I found such pre-estimation unnecessary, as simply taking several thousand additional draws takes a lot less time than writing additional estimation code. In estimation, I therefore follow a conservative approach and initiate the MCMC sampler at an arbitrary starting point (i.e., setting regression coefficients to zero and unobserved heterogeneity variances to 0.1). I discard the first $T = 5,000$ draws and use the next $T^* = 5,000$ draws to describe the posterior distribution of structural parameters and of key counterfactual quantities.

D MCMC Sampler: Simulation Study

To test the performance of the proposed MCMC sampler, I construct the following artificial dataset. I generate a sample of $N = 500$ users in a market with $J = 10$ items and an outside option. Each user does their shopping on week t, which is a uniform random draw from a set of 52 weeks. Items are characterized by two attributes: price p_{jt} and time-invariant attribute x_j. I generate the price of item j in week t as $p_{jt} \sim N(0, 1)$ i.i.d. across weeks and items, and I generate the attributes $x_j \sim N(0, 1)$ i.i.d. across items and independent from prices.

One practical consideration is how to choose the true values of parameters for this simulation study. I could, of course, show that the estimation method "works" for some arbitrary parameter values, but doing so would be of little use for verifying that the MCMC sampler can successfully recover parameters in my application. Instead, I construct a dataset that looks similar to the actual dataset. I set parameter values that generate search behavior similar to that observed in the data. Specifically, I choose parameter values to ensure that (a) most users search 1-2 items, (b) the average number of searches per user is around 2.0, and (c) the majority of users select the outside option. I assume that the utility of user i for item j is given by:

$$u_{ij} = \beta_{1i} + \beta_{2i} \cdot p_{j,t(i)} + \beta_{3i} \cdot x_j + \eta_{ij} + \varepsilon_{ij} \qquad with \qquad \sigma_\eta = 0.5$$

where the heterogeneity of preferences $(\beta_{1i}, \beta_{2i}, \beta_{3i})$ and search propensities ξ_i is given by the following distributions:

$$\begin{aligned}
\beta_{1i} =& \beta_1 & \beta_1 = -3.0 \\
\beta_{2i} =& \beta_2 + \pi_{21} \cdot w_i^p + \sigma_2 \cdot \nu_{2i} & \beta_2 = 0.5,\ \pi_{21} = 0.5,\ \sigma_2 = 0.1 \\
\beta_{3i} =& \beta_3 + \pi_{31} \cdot w_i^p + \sigma_3 \cdot \nu_{3i} & \beta_3 = 0.5,\ \pi_{31} = 0.5,\ \sigma_3 = 0.1 \\
\xi_i =& \xi + \pi_c \cdot w_i^s + \sigma_c \cdot \nu_{ci} & \xi = 2.0,\ \pi_c = 0.2,\ \sigma_c = 0.1
\end{aligned}$$

where ν_{2i}, ν_{3i}, and ν_{ci} are standard normal variables distributed i.i.d. across users and independently from each other.

I emphasize several aspects of this specification. First, I exclude search cost shifters w_i^s from taste parameters β_i, and I exclude taste shifters from the search cost propensity ξ_i. This mimics my empirical application in which I use both preference and search cost shifters to aid identification (see Section 3.4. for a detailed discussion of identification). Removing these exclusion restrictions makes it a lot harder to identify preference parameters separately from search costs. Second, I assume that all users have identical preferences β_{1i} for buying the inside goods. While preference heterogeneity in β_{2i} and β_{3i} can be identified from observed search sequences (see Section 3.4.), I do not observe any data moments that would identify the heterogeneity in how users value the inside goods relative to the outside option. Given the assumed parameter values, 50% of users search only one item, the average user searches 2.0 items, over 50% of users choose the outside option, and all items have non-zero market shares that vary between 0.2% and 7.7%.

Having constructed the artificial data sample, I estimate parameters using the MCMC sampler in Section 3.3. An appealing property of this estimator is its numerical stability. The sampler converges to the same posterior distributions of parameters regardless of the initial point from which I initiate the chain. I therefore initiate the sampler from an arbitrary starting point, let the MCMC sampler run for 5,000 iterations (as in estimation with real data in Section 4) for the burn-in period during which the chain is learning to efficiently traverse the relevant posterior distribution, and I use the subsequent 5,000 draws for inference while discarding the burn-in draws. For this simulation study, I report means and standard deviations of posterior distributions, although in actual applications one would, of course, need to examine the entire distribution of posterior draws for valid inference. I use all 5,000 main draws to compute posterior means, and I report conservative standard errors using a thinned chain that retains only each 100th draw, as suggested by Link and Eaton (2012).

Figure 8 plots the distributions of posterior draws constructed from the last 5,000 draws of the chain. As these plots show, the posterior distributions are centered on values close to the true parameter values. Table 10 reports the corresponding posterior means and standard deviations, showing that most parameters are precisely estimated and have the posterior means close to the true values of parameters. Table 11 shows that the model fits the data well both in-sample and out-of-sample. Specifically, the model can accurately predict the average attributes of searched and purchased products, average search intensity, and the outside option share.

Although the proposed estimator can recover the true parameter values, my final goal is to study consumer surplus from new products. I therefore need to test also whether the proposed estimation method successfully recovers the counterfactual levels of consumer surplus. To this end, I randomly remove 50% of items from the choice set, thus imitating my application where I remove all SSDs from the assortment of hard drives on Amazon. I then measure how this change affects the expected consumer surplus, repeating this computation twice: for the true and the estimated parameter values. To quantify the uncertainty in the predicted consumer surplus, I compute the surplus change for each of the 5,000 posterior draws and plot the resulting distribution in Figure 9. This figure plots the change in consumer surplus from expanding the choice set, which is why the distribution covers only positive values. As Figure 9 shows, the resulting posterior distribution has low variance and is centered on the true value of the consumer surplus change. Therefore, I conclude that the model can indeed accurately recover the relevant welfare estimates.

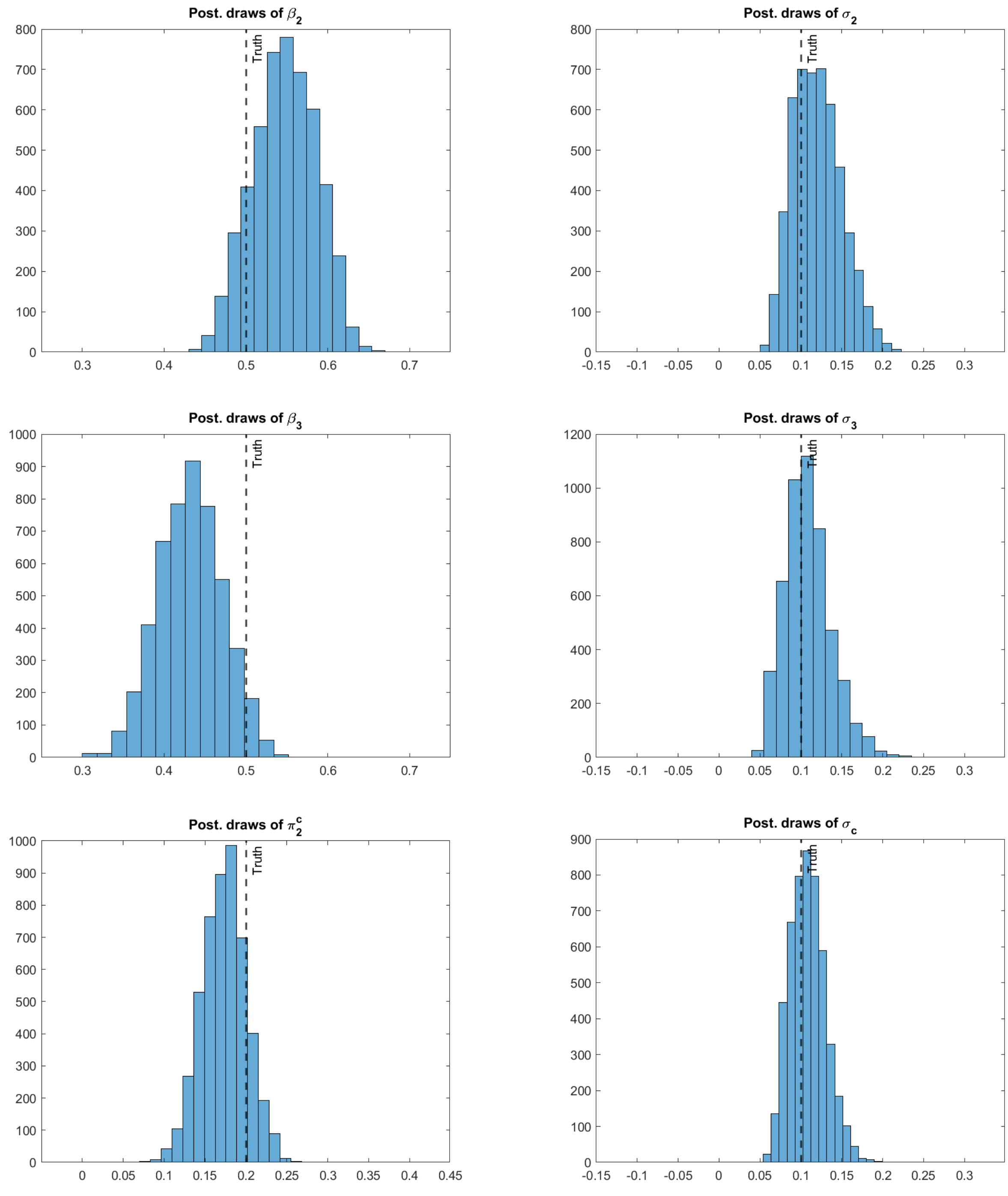

Figure 8: **Posterior distributions from the simulation exercise.** The histograms show the posterior distributions of the model's parameters. The dashed lines show the true (assumed) parameter values.

Parameter	Coef.	Initial Point	True Value	Posterior Mean	Posterior S.E.
Inside option	β_1	0.000	-3.000	-2.838	(0.067)
Price coefficient	β_2	0.000	0.500	0.548	(0.043)
Preference for x_j	β_3	0.000	0.500	0.432	(0.039)
Coef. of β_{2i} on w_i^p	π_{21}	0.000	0.500	0.484	(0.038)
Coef. of β_{3i} on w_i^p	π_{31}	0.000	0.500	0.417	(0.038)
Unobs. heterog. of β_{2i}	σ_2	0.100	0.100	0.120	(0.029)
Unobs. heterog. of β_{3i}	σ_3	0.100	0.100	0.109	(0.031)
Search propensity	ξ_c	2.500	2.000	1.938	(0.047)
Coef. of ξ_{ci} on w_i^p	π_c	0.000	0.200	0.173	(0.029)
Unobs. heterog. in ξ_c	σ_c	0.100	0.100	0.108	(0.029)
Standard dev. of η_{ij}	σ_η	1.000	0.500	0.466	(0.089)

Table 10: **Parameter estimates from the simulation exercise.**

Moment	Training Sample		Test Sample	
	Simulated	Data	Simulated	Data
Avg. x_2 for searched items	0.807	0.762	0.762	0.745
Avg. x_3 for searched items	0.760	0.766	0.759	0.756
Avg. x_2 for purchased items	1.196	1.216	1.169	1.285
Avg. x_3 for purchased items	0.896	0.948	0.947	0.904
Outside option share	0.800	0.780	0.776	0.756
S.E. of x_2 (first searches)	1.039	1.006	1.100	1.144
S.E. of x_3 (first searches)	0.494	0.519	0.491	0.532
Avg. number of searches	2.164	2.192	2.104	2.044
Searched 1 item	0.440	0.428	0.480	0.488
Searched 2 items	0.252	0.240	0.204	0.240
Searched 3 items	0.140	0.164	0.156	0.124
Searched 4 items	0.092	0.084	0.088	0.084
Searched 5+ items	0.076	0.084	0.072	0.064

Table 11: **Model fit from the simulation exercise.**

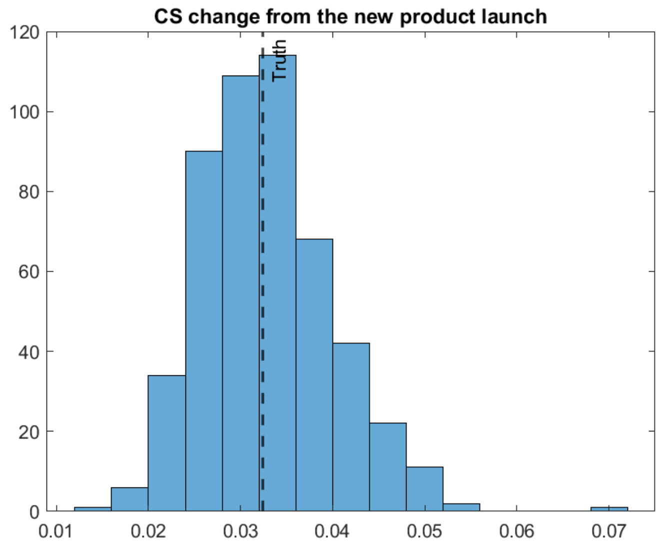

Figure 9: **Welfare estimates from the simulation exercise.** The histogram depicts the posterior distribution of the surplus change from removing 50% of items from the choice set. The dashed line shows the true surplus change given the search model and the assumed parameter values.

E Additional Analyses

E.1 Computing marginal data density

To compare model fit across specifications, I use the implied marginal data density as the main measure of fit. The marginal data density is defined as $p(y|M) = \int p(y|\theta)p(\theta)d\theta$, where y is data, M is the assumed choice model, θ is a vector of structural parameters, $p(y|\theta)$ is the conditional likelihood of data, and $p(\theta)$ is the prior. Following Chib (1995), I compute the log marginal density $\log p(y|M)$ by approximating it with the posterior draws produced from my MCMC sampler. Using the fact that $p(y|M)$ is the normalizing constant in the posterior density, I first apply the Bayes' rule to write the log marginal density $\log p(y|M)$ as follows:

$$\log p(y|M) = \log f(y|\theta^*) + \log \pi(\theta^*) - \log \pi(\theta^*|y) \tag{6}$$

This identity holds for any parameter value θ, but for efficiency reasons, I choose $\theta = \theta^*$ to be a high-density point, which makes it easier to obtain efficient estimates of relevant densities. In practice, I set θ^* to be equal to the posterior mode computed based on the MCMC draws. Computing (6) is relatively easy. The first term, $\log f(y|\theta^*)$, is the conditional data likelihood evaluated at θ^*. Since I only need to compute this likelihood once, I can approximate it with a frequency estimator $\log \hat{f}(y|\theta^*)$ that uses an extremely large number of draws. The second term, $\log \pi(\theta^*)$, is the log density of the prior defined in a closed-form in Appendix C. The last term, $\log \pi(\theta^*|y)$, can be computed by leveraging the fact that my MCMC sampler relies on data augmentation. Specifically, following Chib (1995, p.1314). I re-write the density $\log \pi(\theta^*|y)$ as an integral over the latent data z:

$$\pi(\theta^*|y) = \int \pi(\theta|y, z)p(z|y)dz \tag{7}$$

In my model, latent data z include pre-search utilities δ_{ij}, post-search utilities u_{ij}, consideration indices q_{ij}, and user types λ_i. Since my MCMC sampler produces a sequence of draws for both θ and z, denoted as $\{\theta^{(g)}, z^{(g)}\}_{g=1}^{G}$, I can approximate $\pi(\theta^*|y)$ with a consistent Monte Carlo estimate $\hat{\pi}(\theta^*|y) = (1/G)\sum_{g=1}^{G} \pi(\theta^*|y, z^{(g)})$, which converges to $\pi(\theta^*|y)$ as G becomes large. For any given draw of latent data $z^{(g)}$, I can compute the conditional density $\pi(\theta^*|y, z^{(g)})$ using the same conditional distributions as used by the MCMC sampler (derived in Appendix C). Using these results, I approximate the log marginal density in (6) using the following estimate:

$$\log \hat{p}(y|M) = \log \hat{f}(y|\theta^*) + \log \pi(\theta^*) - \log \left(\frac{1}{G} \sum_{g=1}^{G} \pi(\theta^*|y, z^{(g)}) \right) \tag{8}$$

E.2 Alternative utility specifications

Table 12 explores the robustness of my results in Section 4.1. In the first row, I reproduce the estimated surplus change ΔCS from the main specification in Section 4.1 and compute the log marginal density as a measure of fit (see Appendix E.1). In the second row, I make indirect utility u_{ij} a non-linear function of attributes x_j. I split each continuous attribute into several bins of approximately similar sizes, thus discretizing storage capacity (bins: less than 1TB, 1-2TB, more than 3TB) and speed (bins: less than 100MBs, 100-200MBs, 200-500MBs, more than 500MBs). I then include indicators that a given hard drive belongs to a specific bin in the vector of attributes x_j. In the third row, I maintain the functional form assumptions of the main specification but include additional attributes collected from userbenchmark.com and amazon.com. These attributes include: (a) the Amazon rating displayed on product pages, (b) the overall performance rating of hard drives from userbenchmark.com, and (c) the variance of reported benchmarks on userbenchmark.com as a measure of reliability. As Table 12 shows, none of these alternative specifications generate substantially better fit than the main specification from Section 4.1. I obtain comparable welfare estimates across these specifications. Although changing the functional form of the utility function somewhat increases the estimated surplus from SSDs (specifications in rows 2-3), all four estimates of ΔCS remain higher than the estimate from the perfect information model (see Section 4.3). In other words, changing the model specification does not affect my main qualitative conclusions regarding the bias in the estimated consumer surplus from SSDs.

E.3 The role of rankings

My identification argument in Section 3.4 assumes price changes do not affect consideration. One may worry that Amazon encourages users to consider specific hard drives by placing them on the first page of the hard drive category. To examine the impact of Amazon rankings, and to understand whether accounting for rankings affects my main results, I have collected data on product rankings on Amazon from the Wayback Machine. The archives of the Wayback Machine allow me to reconstruct the appearance of Amazon's pages on different dates. Although the archive does not contain all pages from the hard drive category and does not have data for every single day, I was able to retrieve the identities of all hard drives ranked on the first page on 17 different dates in 2016. I imputed rankings on all other dates using data from the closest available date. For each website visit, I then computed what percent of 25 hard drives presented on the first page were SSDs on that specific date. I then included the resulting "SSD ranking share" into the model as a shifter of consideration parameters γ_i^{SSD} and γ_i^{HDD}. I found that rankings do indeed affect consideration, but the effect is relatively small. Increasing the share of SSDs on the first page by 10 percentage points only increases the consideration probability of SSDs by about 2%. As shown in Table 12 (fourth row), including rankings makes model fit marginally worse and leads to a very similar estimate of the surplus change ΔCS than in the full model without rankings. Although, unfortunately, I do not have higher-quality ranking data, based on these results I do not expect that accounting for Amazon rankings would dramatically impact my welfare results.

Model	Estimated ΔCS from SSDs	Model fit Log Density $\log P(y)$
1. Main specification (estimated in Section 4)	*\$3.23*	*-3013.34*
2. Attributes x_j enter utility non-linearly	\$3.69	*-3013.77*
3. Additional attributes (rating, reliability)	\$3.72	*-3014.01*
4. Amazon rankings entering consideration	\$3.27	*-3013.96*

Table 12: **Robustness analyses.** The first row reproduces estimates from the main specification in Section 4.1. The second row extends the main specification by including speed and storage capacity nonparametrically into the utility function. The third row extends utility to include additional attributes collected from userbenchmark.com and amazon.com. Column 1 reports the estimates of the surplus change from SSDs defined and discussed in Section 4.3. The last column reports the log marginal data density $\log P(y)$ defined in Appendix E.1.

Model	Estimated ΔCS from SSDs	Model fit Log Density $\log P(y)$
1. Main specification (estimated in Section 4)	*\$3.23*	*-3013.34*
2. Search only (consider all items)	\$2.25	*-3018.81*
3. Consideration only (zero search cost)	\$1.41	N/A*
4. Full model w/o heterogeneity (homog. types λ_i)	\$6.00	*-3020.03*

Table 13: **Welfare estimates and out-of-sample fit from alternative models with information frictions.** The first row shows the estimates from the full model with category consideration and costly search developed in Section 3. The second row estimates the same model but shuts down category consideration by assuming all users consider all hard drives. The third row estimates a model with category consideration but omits search by setting all users' search costs to zero. The fourth row estimates the full model but assumes homogeneous preferences, search costs, and consideration parameters. The last column reports the log marginal data density $\log P(y)$ defined in Appendix E.1. *Since the model without costly search by definition cannot predict search behavior, I do not compute the measure of fit for this specification.

E.4 Alternative information models

One question posed in Section 4.2 was whether modeling both category consideration and costly search is necessary for explaining search and purchase data. To further explore this question, in Table 13, I compare the results from the main model with two alternative specifications: one that only models sequential search but omits category consideration (i.e., "search only" model), and another specification which models category consideration but omits costly search (i.e., "consideration only" model). Table 13 presents the results from these two alternative specifications (rows 2-3) and compares them to those in the main specification (row 1).

I find that both models underestimate consumer surplus from SSDs by approximately 40-60%, returning the estimates of \$2.25 (search only model) and \$1.41 (consideration only model). This bias arises because both models underestimate the impact of information frictions on choices. As a result, both models erroneously interpret rare SSD searches and purchases as a sign that users do not like the combinations of attributes offered by the SSDs. A model without search, for example, predicts that mean preferences $\bar{\beta}$ for hard drive speed and Samsung brand are both negative, in contrast to positive estimates from the full model with category consideration and costly search (Section 4.1). Both models return biased estimates of users' preferences because they assume users know more about HDDs and SSDs than they really do (De Los Santos et al., 2012; Goeree, 2008). As a whole, these results suggest that to estimate consumers' benefits from SSDs, one needs to explicitly account for both category consideration and costly search.

Finally, note that Table 13 also shows the welfare results from a model without heterogeneity, observed or unobserved. This model assumes that all three sets of parameters – preferences, search costs, and consideration parameters – are homogeneous across users, so that $\sigma_k^2 = 0$ and $\pi_k = 0$ for all elements k of the type vector λ_i. As expected, the model exhibits worse fit, mostly because it cannot explain why different users focus on different attributes when searching. It also overstates the surplus change from SSDs, generating an estimate of \$6.0, twice higher than in the main specification. I therefore conclude that modeling consumer heterogeneity is crucial for obtaining plausible welfare estimates.

F Model Fit

To validate the estimated model, I explore its ability to predict key data moments out-of-sample. To this end, I split the main dataset into two parts, holding out a random subsample of 50% of users as a test sample ($N = 926$), and using the remaining 50% of consumers as a training sample for model estimation ($N = 926$). In this validation exercise, I focus on four types of moments. First, I predict average attributes of searched and purchased hard drives, which allows me to verify that the estimates of average tastes $\bar{\alpha}$ and $\bar{\beta}$ are meaningful. Second, I predict the variance of attributes among users' first searches to assess the estimates of observed and unobserved heterogeneity. When the "true" tastes are heterogeneous, each user starts their search with a hard drive whose attributes x_j look ex-ante most attractive. The variance of attributes x_j in the first searches then indicates the degree of taste heterogeneity. Third, I compute the predicted distribution of search set sizes to assess whether the search behavior implied by the model resembles that in the data. Finally, I also predict the share of all searches and purchases allocated to HDDs and to SSDs, the moments that directly inform welfare counterfactuals.

Table 14 illustrates fit of the estimated full model from Section 4. The model performs well at predicting key moments, both in-sample and out-of-sample. It performs well at predicting attributes x_j of searched and purchased hard drives, suggesting reasonable estimates of average tastes. It also does a good job explaining the variance of attributes x_j of first searches. This good fit is mainly driven by the estimated (observed and unobserved) preference heterogeneity; in fact, artificially shutting down taste heterogeneity in the estimated model generates 30-50% lower variances of first searches than those observed in the test sample. The model does somewhat worse in fitting the distribution of search set sizes. In particular, while it predicts the average number of searches well, it somewhat overestimates out-of-sample the number of users who searched only one hard drive and the number of users who searched five and more hard drives. This finding suggests that the model with normally distributed search propensities ξ_i may not be flexible enough to capture the underlying distribution of search costs c_i, although I did not attempt to relax the normality assumption for computational reasons. Finally, I also find that the model performs well at predicting how often users search and purchase SSDs compared to HDDs. This is especially reassuring given that in my counterfactuals, I need to predict out-of-sample how many users purchase SSDs once this new subcategory is introduced.

Moment	Train. Sample ($N = 926$)		Test Sample ($N = 926$)	
	Simulated	Data	Simulated	Data
Avg. attributes of searched items $E(x_j \mid j \in S_i)$				
Price *(100s of dollars)*	0.939	0.909	0.930	0.907
Storage 1-2 TB	0.689	0.705	0.688	0.713
Storage 3+ TB	0.158	0.140	0.159	0.132
Speed (100s of MB/s)	1.314	1.293	1.316	1.278
Internal Drive	0.420	0.451	0.417	0.429
Seagate Brand	0.393	0.403	0.403	0.372
WD Brand	0.337	0.332	0.338	0.351
Samsung Brand	0.109	0.100	0.111	0.105
Avg. attributes of purchased items $E(x_j \mid y_i = j)$				
Price *(100s of dollars)*	1.028	0.846	0.957	0.863
Storage 1-2 TB	0.643	0.622	0.623	0.612
Storage 3+ TB	0.226	0.176	0.227	0.149
Speed (100s of MB/s)	1.341	1.151	1.349	1.204
Internal Drive	0.371	0.387	0.386	0.366
Seagate Brand	0.416	0.420	0.441	0.343
WD Brand	0.276	0.303	0.295	0.276
Samsung Brand	0.158	0.050	0.155	0.112
Std. of attributes x_j among first searches				
Price *(100s of dollars)*	0.491	0.555	0.474	0.539
Storage 1-2 TB	0.464	0.468	0.464	0.467
Storage 3+ TB	0.357	0.358	0.361	0.356
Speed (100s of MB/s)	1.158	1.210	1.162	1.214
Internal Drive	0.493	0.499	0.493	0.494
Seagate Brand	0.488	0.487	0.487	0.481
WD Brand	0.468	0.466	0.472	0.471
Samsung Brand	0.320	0.306	0.324	0.319
Outside option share	0.761	0.871	0.762	0.855
Avg. number of searches	1.661	1.551	1.603	1.552
Searched 1 item	0.690	0.684	0.706	0.676
Searched 2 items	0.150	0.193	0.145	0.200
Searched 3 items	0.078	0.072	0.068	0.075
Searched 4 items	0.026	0.026	0.041	0.029
Searched 5+ items	0.056	0.025	0.040	0.021
Share of searches SSDs	0.166	0.170	0.164	0.175
Share of searches HDDs	0.834	0.830	0.836	0.825
Share of purchases of SSDs	0.140	0.134	0.164	0.172
Share of purchases of HDDs	0.860	0.866	0.836	0.828

Table 14: **Estimated model fits the key data moments.** The table compares moments predicted by the estimated model to those observed in the data, both for a training sample used for estimation ($N = 926$) and a test sample used for out-of-sample validation ($N = 926$).

G Computing Consumer Surplus

This section explains how I compute consumer surplus in Section 4.3. I define consumer surplus as the expected value of ex-ante utility, where the expectation is taken with respect to user types λ_i, consideration sets C_i, search sets S_i, and purchase decisionsy_i. Specifically, I compute consumer surplus as:

$$CS(J) = \int (u_{i,y_i} - |S_i - 1| \cdot c_i) dF(y_i, S_i, C_i|\lambda_i) dF(\lambda_i|\rho) \tag{9}$$

where J denotes the set of hard drives in the current assortment (e.g., $J = SSD$ in the counterfactual scenario), u_{i,y_i} is the utility of the purchased hard drive y_i, $|S_i - 1|$ is the number of times the user incurred the search cost c_i (i.e., recall the first search is free), and ρ denotes hyperparameters that determine the distribution of types $F(\lambda_i|\rho)$.

In practice, computing CS in (9) takes a lot of time and suffers from approximation errors. To circumvent this issue, I follow the result of Choi et al. (2018) who show that a sequential search model can be reformulated as a discrete choice model with modified indirect utilities. According to this result, my model is equivalent to a discrete choice model in which users choose an option with the highest *effective utility* v_{ij} defined as $v_{ij} = \min(u_{ij}, z_{ij})$ where the reservation utilities are given by $z_{ij} = \delta_{ij} + \xi_i$. This alternative formulation removes the need to repeatedly solve the optimal search problem, thus simplifying welfare computations. I then compute consumer surplus from the choice set J as follows:

$$CS(J) = E\left(\max_{j \in C_i}\{v_{ij}(\theta_i, c_i)\}\right) = \int \max_{j \in C_i}\{v_{ij}(\theta_i, c_i)\} dF(C_i|\lambda_i) dF(\lambda_i|\rho) \tag{10}$$

where $\max_{j \in C_i}\{v_{ij}(\theta_i, c_i)\}$ captures the highest effective utility among hard drives in the consideration set C_i, and $F(C_i|\lambda_i)$ denotes the distribution over potential consideration sets $C_i \subseteq J$ given the type λ_i. To compute the surplus via simulation, I then simply draw user types λ_i and consideration sets C_i, as well as item-specific shocks η_{ij} and ε_{ij}, compute the maximum of the resulting effective utilities $v_{ij}(\theta_i, c_i)$, and average across simulations.[18] In practice, I draw $100,000$ users (i.e., types λ_i) and then draw 10 sets of random shocks for each of these simulated users, averaging the resulting maximum of effective utilities across draws. Further increasing the number of draws only marginally changes the surplus estimates but substantially increases the computation time. Having computed the surplus estimates, I then calculate the desired surplus change as the difference $\Delta CS = CS(HDD \cup SSD) - CS(HDD)$.

[18] Note that the formulation in Choi et al. (2018) assumes that the first search is costly, whereas in my model the first search is free (see Section 3.2). However, in practice, accounting for the first search barely makes any difference for welfare estimates, which I verified by comparing the approximations of (9) and (10) that use very large numbers of draws. In general, formula in (10) gives a sufficiently precise estimate of the expected consumer surplus.